

Impact of next generation sequencing techniques in Food Microbiology

Baltasar Mayo¹, Caio T. C. C. Rachid², Ángel Alegría¹, Analy M. O. Leite², Raquel S. Peixoto², and Susana Delgado¹

Departamento de Microbiología y Bioquímica, Instituto de Productos Lácteos de Asturias (IPLA-CSIC), Paseo Río Linares s/n, 33300-Villaviciosa, Asturias, Spain¹, and Instituto de Microbiologia, Departamento de Microbiologia Geral, Universidade Federal do Rio de Janeiro, Avenida Carlos Chagas Filho, 373, 21941904-Cidade Universitária, Rio de Janeiro, RJ, Brazil²

Keywords: Next generation sequencing, NGS techniques, pyrosequencing, Illumina, Food Microbiology, Molecular Microbial Ecology, Food Ecology

Short title: NGS techniques in Food Microbiology

*Correspondence to:

Baltasar Mayo, Departamento de Microbiología y Bioquímica, Instituto de Productos Lácteos de Asturias (IPLA-CSIC), Paseo Río Linares s/n, 33300-Villaviciosa, Asturias, Spain

Phone number: +34985892131

Fax number: +34985892233

E-mail: baltasar.mayo@ipla.csic.es

Abstract

In recent years there has been an explosion of next generation sequencing (NGS) techniques (understanding the Maxam-Gilbert and Sanger techniques as the first generation). NGS techniques have high-throughputs and produce thousands or even millions of sequences at the same time. These sequences allow for the accurate identification of microbial taxa, including uncultivable organisms and those present in small numbers, and provides a complete inventory of all microbial operons and genes present or being expressed under different study conditions. NGS techniques are revolutionizing the field of microbial ecology and have recently been used to examine food ecosystems. After a short introduction to the most common NSG systems and platforms, this review addresses how NGS techniques have been employed in the study of food microbiota and food fermentations, and discusses their limits and perspectives. The most important findings are reviewed, including those made in the study of the microbiota of milk, fermented dairy products, and plant-, meat- and fish-derived fermented goods. The knowledge that can be gained on microbial diversity, population structure and population dynamics via the use of these technologies could be vital in improving the monitoring and manipulation of foods and fermented food products. They should also improve their safety.

1. General introduction

Foods harbor complex microbial communities composed of viruses, bacteria and fungi. Some of these microorganisms are of technological importance and invest foods with desirable sensorial (organoleptic and rheological) characteristics [1]. However,

undesirable microorganisms may also be present; these can reduce the quality of foods (spoilage microorganisms) or even negatively affect their safety (pathogens) [2].

The traditional way of determining the composition of food-associated microbiota has been through culturing methods. These rely on the isolation and cultivation of microorganisms before their identification and typing. However, it has been repeatedly shown that culturing is unreliable for the complete microbial characterization of many ecosystems, including those of foods [3]. The selective isolation of microorganisms may require unknown growth factors and/or growth conditions present in natural habitats that are not reproduced by laboratory media. Foods may also have a low pH, a reduced a_w , or have to be kept under harsh storage conditions, etc. which might leave microbes in a physiologically viable but not cultivable state. Moreover, microbes present in low numbers can be outcompeted by numerically more abundant species, impeding their detection in culture. Such limitations lead to these techniques underestimating microbial diversity, and sometimes even the failure to detect the majority microbial groups [4]. Over recent decades, a great number of culture-independent, molecular methods have been developed that help overcome these problems, most of which have been used extensively in food systems (for reviews see [5, 6]). These techniques allow the identification and quantification of food-associated microbial groups, and provide sensitive and rapid methods for determining the composition and diversity of complex microbial communities.

Culture-independent techniques are mostly based on the analysis of microbial nucleic acid sequences (DNA and/or RNA). Most rely on the amplification of these by the polymerase chain reaction (PCR). The majority of amplification techniques target the 16S rRNA gene (rDNA). Comparing the sequences obtained to one another, and to

those held in databases, allows phylogenetic relationships between microbes to be established. Culture-independent, PCR-based techniques include denaturing gradient gel electrophoresis (DGGE), temporal temperature gradient gel electrophoresis (TTGE), single stranded conformation polymorphism (SSCP), real-time quantitative PCR (qPCR), the construction and analysis of 16S rRNA gene libraries, and a few others. Of these, DGGE (qualitative analysis) and qPCR (quantitative analysis) have been widely used to microbiologically characterize food environments and to analyze the course of dairy fermentations [2, 6]. Changes in microbial populations furnish useful information regarding the dynamics of food fermentation, the monitoring of the growth of starter and adjunct cultures, and the fate of spoilage microorganisms and pathogens. A number of new microbial types with no cultured relatives – but of potential technological interest – have also been detected using these techniques [7-14].

NGS techniques have promoted the emergence of new, high-throughput technologies, such as genomics, metagenomics, transcriptomics and metatranscriptomics, etc. These technologies can be used in two substantially different ways: sequencing the total microbial nucleic acids (shotgun sequencing) and gene-specific sequencing (targeted sequencing). For the latter, segments of highly conserved DNA or cDNA sequences are first amplified by PCR using universal or group-specific primers. Shotgun sequencing returns information far beyond that regarding phylogenetic composition, providing insights into the number and potential function of genes within the community [15-16]. Both shotgun and targeted techniques have already been used to study the microbiology of a series of foods and food fermentations, and pertinent reviews have been compiled [16-19]. However, research in this area is so

active that findings must be continually reviewed, and the current and potential applications of these constantly updated.

2. NGS platforms

NGS platforms involve many different technologies [20] all of which generate large, genome-scale datasets. However, they differ substantially in terms of their engineering, sequencing chemistry, output (length of reads, number of sequences), accuracy and cost. Current commercial platforms include the Roche 454 (Roche), Solexa (Illumina), SOLiD and Ion Torrent (Life Technologies), and PacBio (Pacific Biosciences) systems. Comparisons of their advantages and disadvantages have recently been published [21, 22]. Many other ‘third-generation’ techniques, such as DNA nanoball sequencing, heliscope single molecule sequencing, nanopore DNA sequencing, tunneling current DNA sequencing, sequencing with mass spectrometry, and microscopy-based techniques, are currently under development [23].

2.1. The Roche 454 pyrosequencer

The NGS era started with the release of the first array-based 454 pyrosequencing system by Roche in 2004 [24]. In this system, double stranded DNA fragments are ligated to specific adaptors, diluted, and immobilized on tiny beads (one molecule per bead). The attached DNA is copied millions of times by PCR, with one of the primers biotinylated, in an oil-water emulsion. The beads are then housed in separate picotiter-size wells, in which the non-biotinylated strand is denatured and washed away. The plate array consists of approximately one million wells, and independent sequencing reactions occur within each. After the addition of the sequencing primer, the

pyrosequencing reaction proceeds via DNA synthesis using single stranded DNA molecules as templates [25]. The sequencer runs the A, T, C and G nucleotide building blocks – one type at a time - over the wells. When a nucleotide is incorporated, pyrophosphate is released and converted into ATP, which fuels a coupled luciferin-luciferease reaction producing light (the intensity proportional to the number of incorporated nucleotides). Light is recorded by a camera, and the unincorporated nucleotides degraded enzymatically. At present, 454 pyrosequencers offer reads longer than 1000 bp, with a high sequence output (1 megabase (Mb) in the case of the 454 GS FLX+ system). Difficulties in sequencing homopolymer stretches and the rather high cost per Mb are the main weaknesses of the 454 platform.

2.1. The Solexa platform

The technology used by this platform was released by Solexa in 2006, a company later purchased by Illumina. In this method, primers with specific adaptors are attached to a slide and the DNA molecules amplified by PCR so that local clonal DNA clusters are formed. Sequencing proceeds by synthesis, using four types of reversibly-blocked terminator nucleotides (RT-nucleotides). Non-incorporated RT-nucleotides are washed away. The dye, along with the terminal 3' blocker, is then chemically removed from the DNA, allowing for incorporation detection and for the next sequencing cycle to begin [26]. Unlike in pyrosequencing, DNA chains are extended one nucleotide at a time. Additionally, image acquisition can be delayed until after nucleotide incorporation, allowing for very large arrays of DNA clusters to be captured by sequential images taken by a single camera. Currently, the Solexa platform generates typical reads of 150-300 bp that can be increased to 300-600 bp via “paired-end” sequencing (i.e.,

sequencing both ends of the same DNA segment). The reduced price for the high output of sequences per run (up to 3000 Mb for the HiSeq) is the greatest advantage of this technology.

2.3. The SOLiD system

The first SOLiD (Sequencing by Oligonucleotide Ligation and Detection) sequencer was released in 2007 by Life Technologies. True to its name, and unlike any other technology, the SOLiD technique carries out sequencing by ligation instead of synthesis [27]. DNA fragments are ligated to a universal adapter, and then attached to a magnetic bead. After emulsion PCR, the resulting amplicons attached to the beads are bound covalently to a glass slide. A set of four fluorescently-labeled di-nucleotide probes compete for ligation to the sequencing primer (complementary to the adaptor). Following a series of cycles of probe ligation, dye cleavage and detection, the extension product is removed and the template reset with a primer complementary to the n-1 position for a second round of ligation. Sequencing by ligation is very accurate (99.99%) since ligases make fewer mistakes than polymerases. However, the maximum-length reads obtained with SOLiD measure just 75 bp, insufficient for most metagenomic studies. This also makes the assembly of the sequences a very challenging task. SOLiD has uses in whole genome resequencing, targeted sequencing, and transcriptome research.

2.4. Ion Torrent

The Ion Torrent Company (now acquired by Life Technologies) released its first Personal Genome Machine (PGM) sequencer based on semiconductor technology in

2010. As in other methods, sequencing is performed via the synthesis of a complementary strand, but detection is based on the hydrogen ions released during DNA polymerization; the number of releases is proportional to the number of incorporated nucleotides [28]. This sequencing technology differs from others in that no modified nucleotides or optics are used. After library construction and emulsion PCR, microwells containing single template pools of the DNA to be sequenced are flooded in turn with single nucleotide species. The PGM can generate up to 4 Mb in 7 h at a moderate cost per Mb. The length of the reads has risen in recent years up to 400 bp. This technology is less accurate (98%) than others, but is useful in small studies, such as the sequencing of microbial genomes and targeted metagenomics.

2.5. PacBio SMRT

The PacBio platform, marketed in 2010 by Pacific Biosciences, uses single molecule real-time (SMRT) sequencing technology. As in other methods, sequencing is based on synthesis, using nucleotides labeled with distinct fluorescent dyes, but sequencing proceeds without the need for prior PCR amplification. The single stranded DNA molecules to be sequenced are deposited in zeptoliter wells, where a single polymerase molecule is immobilized [29]. The wells are constructed in such a way that only the fluorescence occurring at the bottom of the well is detected. The fluorescent label is detached from the nucleotide during its incorporation into the DNA strand. The accuracy of single reads is low (~85%), but high accuracy can be achieved (99.999%) by building up consensus sequences based on all the reads for the same molecule. Theoretically, large DNA molecules (up to 30 kb) can be sequenced by this technology. The latest protocols have improved the average read length to over 7 kb. However, the

low output obtained with this platform (0.05 Mb per run) is a major shortcoming, rendering it unsuitable for large sequencing projects. Nevertheless, its high accuracy can be exploited for sequencing difficult DNA regions and for genome scaffolding in metagenomic studies.

3. Examining the microbial ecology of foods by NGS

The first attempts to characterize whole community profiles of traditional fermented foods were performed using rDNA amplification, cloning, sequencing and sequence analysis [12, 30-32]. These methods are now being replaced by high-throughput NGS techniques. Targeted techniques provide a snapshot of the diversity and phylogeny of the different elements making up microbial populations. The term phylobiome has been introduced recently to refer to the phylogenetic information gathered using this approach [33]. In addition, shotgun techniques inform on the genetics and functional capabilities of the microbial constituents of food ecosystems.

3.1. Workflow, pitfalls and care

The NGS analysis of nucleic acids from food samples requires a series of steps, each of which must be optimized: nucleic acid extraction and purification, library construction, sequencing and analysis of the raw sequences using computer software, and the searching of different databases (Figure 1). These steps are discussed below. NGS techniques have further witnessed the appearance and spreading of a myriad of molecular microbiological terms (Table 1), to which food microbiologists were unfamiliar until recently.

3.1.1. Nucleic acid extraction and purification

In microbial studies, the handling and storage of food samples is critical. Rapid processing is always recommendable, and all preparations and conditions that might alter the original proportion of microbial cells should be avoided.

Food systems are almost always heterogeneous. Thus, strategies must be carefully designed to ensure that samples are representative of the different conditions found in the food, and that sufficient homogenization is obtained. Incomplete cell lysis, and therefore inadequate DNA extraction, can certainly provide unrealistic views of a food microbial community. This can be avoided by performing supplementary lysis steps to improve extraction efficiency, e.g., via the combination of enzymatic and mechanical treatments [34]. Further, some food components, such as lipids and proteins, may inhibit PCR amplification. These compounds must be removed during nucleic acid extraction [35]. Extracting the same sample multiple times and making a final pool can further minimize the bias associated with single extractions.

Quality control and nucleic acid quantification are vital in any sample preparation method. Optimum DNA concentrations should be used for the instruments available [34]. Accurate quantification is preferentially achieved via the use of fluorometers [34, 36]. The amount and quality of DNA required from a food sample for total DNA sequencing (shotgun metagenomics) is usually higher than that needed for gene-specific sequencing (targeted metagenomics).

In foods, it is of great importance to distinguish between viable, active and inactive cells. Dead cells may contribute to the DNA pool which will be further processed. The use of NGS techniques with cDNA, produced by reverse transcription from RNA, focuses on active microbial populations, thus avoiding these problems [37, 38]. In

metatranscriptomic studies, interest focuses on the non-ribosomal fraction, which includes mRNA, micro-RNA, tRNA and other non-coding RNAs [34]. However, the majority of total RNA corresponds to rRNA, which may account for >90% of the reads obtained [39, 40]. Indeed, in RNA-based studies, sample preparation remains an active area of investigation; the recovery and enrichment of high quality mRNA, its short half-life, and the selective removal of rRNA, are all challenges that must be adequately met [33, 41, 42].

Alternatively, complex communities can be ‘enriched’ in viable cells via the use of propidium monoazide (PMA) or ethidium monoazide (EMA) prior to DNA extraction. These compounds enter cells with damaged membranes and bind covalently to DNA, preventing subsequent PCR amplification [43]. The effectiveness of these azides has been tested in the microbial analysis of water samples [43].

3.1.2. Molecular target and library construction

Specificity and the coverage of the target sequence are key variables that must be considered in primer design [44]. The target sequence should be long enough to contain nucleotide heterogeneity that distinguishes between organisms. Hypervariable regions within the 16S rRNA gene are commonly chosen for the assessment of taxonomic bacterial and archaeal diversity in foods [45-50]. Primers, however, should be designed in well-conserved, universal regions to cover most (if not all) of the microorganisms present in the sample [51-53]. The preferential amplification of sequences from certain microbial types (due to different gene lengths and/or GC contents) is always a danger [44, 54]. Targeting specific microbial groups may require the amplification of specific hypervariable regions [44]. For instance, Bokulich et al. [46] showed the V4 domain to

be more appropriate than the V5 domain for wine ecology studies since it provided greater taxonomic resolution for lactic acid bacteria (LAB) communities. Some closely related species cannot be truly differentiated by 16S rDNA sequences [53, 55]. Single-copy target genes, such as *rpoB*, *recA*, *radA*, *rpoA*, *gyrB* and others have therefore been proposed as alternatives [55-57]. However, the absence of well-represented databases for comparison is an important limiting factor in their widespread use [18]. Primer design should also avoid homology between prokaryotic and eukaryotic sequences. A further problem associated with the amplification of ribosomal genes is the operon copy number, which varies widely across taxa, distorting quantitative diversity estimates [58], as well as the intra-species 16S rRNA gene heterogeneity found in natural populations [59]. Finally, the use of state-of-the-art, high-fidelity polymerases is recommended to reduce error rates. This can also help prevent base-calling errors and chimera formation, common problems associated with PCR [60].

Several molecular target genes have already been used to track fungal communities. The most commonly used have been the 18S and 28S regions of rDNA [61]. Well established primers and databases are described in the literature [62]. However, they cannot be used to determine the phylogeny of fungi at the species level [61, 63]. Therefore, when this level of discrimination is required, primers targeting the ITS regions are more appropriate [63, 64]. However, these sorts of primer do not align with confidence across genetically distant taxa [65].

In the context of NGS, the library is the entire number of fragments to be sequenced. Library construction includes the joining of platform-specific adaptors and sample-specific identifiers to double stranded DNA fragments. A key difference in library construction for total nucleic acid sequencing (shotgun metagenomics) and gene-

specific sequencing (targeted metagenomics) is the latter's need for the prior PCR-amplification of conserved DNA or cDNA sequences. For targeted metagenomics, platform-specific adaptors are designed as part of the PCR primers, resulting in template molecules ready to be sequenced. The multiplexing of several food samples is possible if short sample-specific DNA sequences known as barcodes, tags or multiplex-identifiers (MIDs) are used [66]. These are also designed as part of the PCR primers. Care must be taken, however, since some barcodes can be more efficiently amplified, providing uneven coverage of the different samples [67]. Normalized mixtures of the amplicons should be made prior to sequencing, thus ensuring that each amplicon is adequately represented in the sequencing reaction [34]. In shotgun sequencing, DNA amplification is not necessary (except when only tiny amounts of DNA are available) [68]; amplification of total environmental DNA is obtained via the use of Phi29 DNA polymerase [69]. Since no amplification is required, shotgun metagenomics avoids the inherent bias of PCR-based techniques. The preparation of the shotgun library commonly starts by fragmenting the DNA using mechanical force. An attractive alternative is digestion by restriction enzymes [34]. In either case, platform-specific adaptors are then ligated to the resulting DNA fragments. Barcodes can also be included as part of the adaptors, thus allowing multiplexing in shotgun sequencing projects.

3.1.3. Data analysis

Reads obtained by NGS techniques need to be converted into more user-friendly information that allows comparisons between samples. In targeted metagenomics, this is generally done by clustering the reads into operational taxonomic units (OTUs) (Figure 2) based on the similarities of the reads themselves (similarity-based approach), or by

taxonomic assignment (composition approach method), i.e., classifying the reads at different taxonomic levels (phylum, class, order, family, or genus) based on the similarities of the sequences with those in a database. Both approaches are currently used and produce similar results [70].

Taxonomic assignment has the advantage of providing information about the relationship of the reads with known microbial groups; this helps in ecological and/or functional interpretations. Additionally, it allows for comparisons between different studies, even when sequences come from different regions of the marker gene (rDNA, for example). The ribosomal database (RDP) [71], Silva [72] and Greengenes projects [73] provide updated databases for the small (16S and 18S) and large (23S and 28S) subunits of rDNA sequences. In the absence of a consensus definition of genetic species, as a rule of thumb, rDNA sequences sharing a 95% of identity are considered to belong to the same genus, and sequences sharing 97% or higher identity are considered to belong to the same species. These percentages are considered for the full rRNA gene molecule; therefore, lower resolution could actually be achieved by analyzing shorter reads. When working with other phylogenetic markers (like *rpoB*, *gyrG*, *recA*, etc.) and functional genes (such as those coding for antibiotic resistance and biodegradation activities, etc.) others databases such as the FunGene (<http://fungene.cme.msu.edu>) should be used. Finally, web tools, such as RAST [74] and MG-RAST [75] provide databases designed to help in the interpretation of genomic and metagenomic data respectively. All these databases have user-friendly web-based pipelines and are good options for users with little bioinformatic training. However, all analyses are pre-formatted, do not allow for customization, and provide no good quality controls. Moreover, the vast majority of microbial types have never been characterized or

classified into taxonomic groups. Thus, there is a lack of representative organisms at several taxonomic levels that limits any taxonomic assignment approach [76].

An alternative and very common way to analyze sequencing information is that based on the genetic similarity between reads. Since most of the 16S rRNA genes are unknown, the diversity of the sample can be measured as the number of different OTUs; this is represented by the clusters constructed at a given level of sequence dissimilarity (normally at 1-3% for bacteria). This approach is advantageous when dealing with complex communities in which most of the microorganisms cannot be classified [76]. However, it provides no information about “who is there”, thus limiting the interpretation of the data. Some of the more recent bioinformatic tools that integrate the taxonomic-dependent and -independent approaches, offer several means of customizing analyses, and provide high quality data, which can easily be presented in comprehensive tables and didactic graphs (Figure 2). The most popular software programs are Mothur [77] and QIMME [78], both of which are freely distributed and come with manuals and data analysis examples. These programs are periodically updated, improving analytical options as new concepts arise. Initial bioinformatic concerns were restricted to the verification of primer and barcode sequences, fragment length, and sequence quality. However, other control steps are now required, including noise removal [79], the detection of chimeric sequences, procedures for preclustering, and the exclusion of singletons [80-82]. Another promising method of NGS data analysis relies on the link between community phylogeny and functionality. Based on the known genome content of a few strains, PICRUST software is able to infer the number of genes shared by different microbial taxa. This program provides information on the phenotypic

relationships within a microbial community based solely on its phylogenetic composition [83].

3.2. Composition, structure and dynamics of food-associated microbial populations

The pyrosequencing of tagged 16S rRNA gene amplicons of DNA and cDNA was first used to analyze the community structure and population dynamics of ten samples of “ben-saalga” pear millet slurry (a traditional African fermented product) [84] (Table 2). This allowed 80.7% of the sequences to be attributed to a family and 70% to a genus, but did not permit identification at the species level. The amplification of large numbers of Archaea (4.8%) and Eukarya (43%) sequences was rather surprising since bacteria-specific primers were used. Further, the number of sequences per sample varied widely (from 641 to 13,380), highlighting the under-representation of the diversity of some samples. With the exception of a few proteobacteria, almost all the bacterial sequences were attributed to cultivable bacteria. As expected, the sequences attributed to LAB genera were the most common (ben-saalga is made via lactic acid fermentation). Further, the results obtained from the analysis of cDNA revealed the same microbiome pattern, suggesting that the assignments made through the DNA analysis corresponded to living bacteria. Considerable diversity was found at the beginning of fermentation, whereas at the end only representatives of the Lactobacillaceae (*Pediococcus* and *Lactobacillus*), Leuconostocaceae and Enterococcaceae were found. Nevertheless, different patterns of community development were seen between samples fermented spontaneously, those inoculated by backslopping, and those inoculated with pure LAB cultures. This was attributed to a poor adaptation of the LAB cultures to the slurries (they were selected from culture

collections). Despite these problems, the authors concluded this NGS method to be very promising for the preliminary and/or rapid microbial characterization of food samples.

3.2.1. Asian, soybean- and rice-derived foods

The bacterial community of “meju”, a Korean, soybean-derived, fermented product, was characterized in one of the first food microbial ecology studies involving NGS [85]. Meju is similar to the Japanese “natto” and “miso”, and to the Indonesian “tempeh”, and is the basis of many Asian products in which fermented soybean pastes are mixed with seasonings, vegetables and rice, e.g., “cheonggukjang” [86], “doenjang” [87], and “kochjang” [88]. The microbial populations of all these products have now been examined using NGS techniques. *Bacillus* spp. (*B. subtilis*, *B. licheniformis*, *B. amyloliquefaciens*, and others) seemed to be dominant throughout fermentation, followed by *Enterococcus* spp. (*E. faecium*, *E. faecalis*), *Lactobacillus* spp., and species of other LAB genera (*Leuconostoc*, *Weissella*, *Tetragenococcus*). However, the actual majority species varied widely between samples, and sometimes the number of LAB reads surpassed that of the bacilli [86-88]. Most of the bacterial species had been identified in these products by culture and culture-independent techniques. However, notable discrepancies have been noted between the results obtained by these methods and those provided by pyrosequencing [86].

Another Korean product which has been actively investigated by pyrosequencing is “kimchi” [89, 90], a fermented product that combines vegetables, fermented seafood products (“jeotgal”), seasonings, and a starchy-rich paste (rice, wheat). There are hundreds of kimchi varieties, depending on the region, local area, and the availability of seasonal ingredients. In ten types of kimchi, the succession of microbes was reported to

follow a few general trends [89]. Together with some species of Proteobacteria (*Pseudomonas*, *Enterobacter*), pyrosequencing showed that *Leuconostoc* spp. and *Weissella* spp. dominate the early stages of fermentation, while *Lactobacillus* spp. are dominant at the end [89].

The archaeal and bacterial communities of seven types of jeotgal - as mentioned above, an ingredient of “kimchi”, but also used as a side dish - have also been analyzed by pyrosequencing [45]. Since the raw ingredients are mixed with large amounts of salt, several members of the family Halobacteriaceae (of the archaeal phylum *Euryarchaeota*), such as *Halorubrum* and *Halalkalicoccus*, were commonly detected in most samples. *Lactobacillus* and *Weissella* reads in similar proportions made up the majority bacteria, although Gammaproteobacteria belonging to the genus *Salinivibrio* were predominant in one sample. The authors concluded pyrosequencing to have revealed new phylotypes of both bacteria and archaea that were not detected by DGGE.

The use of different NGS techniques for the microbial fingerprinting of Japanese, rice-based fermented products, such as “narezushi” [91], “nukadoko” [92] and “kaburazushi” [93] has also been reported (Table 1). These products, which combine water, salted and/or vinegar-treated fish and boiled rice in different proportions, are subjected to lactic acid fermentation. Not surprisingly, *Lactobacillus* spp. made up the dominant populations, followed by other LAB genera. Similar results were found during the microbial typing of a Korean rice-based fermented beer, in which Proteobacteria in the raw substrate were shown finally to be replaced by different LAB species [94]. Further, it was shown that *Saccharomyces cerevisiae* become the dominant yeast during the final stages of the fermentation.

3.2.2. Microbial fingerprinting of plant-derived fermentations

Pyrosequencing of both rDNA- and cDNA-derived amplicons has been used to study the microbial ecology of Spanish-style and Greek-style fermented table olives [38]. In this work, the olive surfaces and brines (8-12% NaCl) were sampled independently. Greek-style fermentation involves natural, untreated olives that are directly brined after picking; Spanish-style fermentation includes a previous washing step of the fruits with a dilute NaOH (2-3%) solution. This reduces bitterness through the degradation of oleuropein and polyphenols, and softens the olive pericarp increasing its permeability. The bacterial compositions revealed throughout fermentation by DNA and cDNA analyses were very similar, suggesting that microorganisms were all alive and active. The initial stages of both the Greek and Spanish fermentations were characterized by large numbers of halophilic bacteria (*Chromohalobacter* and *Halomonas* accounting for 50-60% of the total reads), while at the end of the fermentation (90 days), *Lactobacillus* species surpassed all other microorganisms. The Spanish-style fermentation was associated with larger numbers of Enterobacteriaceae, such as *Enterobacter*, *Citrobacter*, *Escherichia* and *Klebsiella*. These differences may account for the distinctive safety and preservation properties of these two forms of fermented olives.

rDNA pyrosequencing has been used to examine the bacterial ecology of rye and wheat sourdough fermentations [49]. With the exception of the Enterobacteriaceae, all other flour-associated contaminants were shown to be completely inhibited during sourdough propagation. Rye sourdough fermentation was dominated by *Weissella* spp., the dominant bacteria in rye flour. In contrast, in wheat sourdoughs, *Lactobacillus sakei*, *Leuconostoc* spp., *Weissella* spp. and *Lactococcus lactis*, fluctuated as the majority species. The rye and wheat flour types initially showed distinctive microbiota,

but over fermentation their complexity was rapidly simplified and a core microbiota composed mainly of LAB was seen in both sourdough fermentations.

Finally, the Illumina platform has been used to track the bacterial populations during fermentation of the barely-based, ale-type American beer [95]. Enterobacteriaceae were shown to be replaced throughout fermentation by LAB species, among which *Pediococcus* spp. constitute a majority. A vast array of minority species, including some not previously associated with this fermentation, were also detected [95].

3.2.3. Microbiology of milk and fermented dairy products

The pyrosequencing of rDNA amplicons has been used to examine the microbiome of cow milk, with particular interest shown in identifying the microbial types involved in mastitis [96]. In most samples, the mastitis-causing pathogens identified by culturing (*Escherichia coli*, *Klebsiella* spp., *Trueperella pyogenes*, *Streptococcus uberis*, *Staphylococcus aureus*) were among the most common detected by pyrosequencing. However, rDNA sequences of bacterial pathogens not previously associated with mastitis, as well as of bacteria never described as pathogens, were occasionally detected.

The communities in kefir grains and kefir beverages of different origin have also been studied by the pyrosequencing of rDNA amplicons [47, 97]. As expected, the dominant microorganisms detected by culturing were also identified by pyrosequencing; these included *Lactobacillus kefirianofaciens*, *Lactobacillus kefir*, *Lactobacillus parakefir*, *Lactobacillus buchneri*, and others. Pyrosequencing also revealed minor bacterial constituents such as species of *Acetobacter* and *Lactococcus* [47, 97]. This

technique further confirmed and that the microbiota of the beverage is distinct to that of the grains [97].

Pyrosequencing has also been used to study the diversity and dynamics of bacterial populations present during the manufacture and ripening of traditional cheeses, including Danish raw milk cheeses [98], the Polish cheese Oscypek [99], artisanal Irish cheeses [100], Latin-style cheeses [101], and water buffalo Mozzarella [102] (Table 2). Different LAB species were shown dominant in all cheeses and at all stages of production, the actual types depending on the cheese technology employed and/or the use (or not) of starter cultures. Several bacterial types never before reported in traditional and artisanal cheeses were identified by pyrosequencing, including *Bifidobacterium* spp. in Oscypek [99], *Faecalibacterium*, *Prevotella*, and *Helcococcus* in Irish cheeses [100], and *Exiguobacterium* in Latin-style cheeses [101]. The importance of these and other minority populations on the sensory profiles of these cheeses and/or their safety remains uncertain.

The pyrosequencing of DNA- and cDNA-derived amplicons has been used to track the fate of starter cultures and inoculated pathogens (*E. coli*, *Listeria innocua*, *Staphylococcus aureus*) in model cheeses made from raw milk [37]. Wide bacterial diversity was seen to decline during ripening when using either DNA- or RNA-extracted samples. *Listeria* and *Staphylococcus* reads were only detected in DNA from cheeses to which they were added, suggesting these bacteria were in an unviable state. In contrast, *E. coli* was detected at different times during ripening via both rDNA and cDNA libraries; pathogenic strains of this species may represent a hazard in the cheese environment.

The characterization of the diversity of thermophilic bacteria in milk following pasteurization and culturing at 42°C for 24 h has also been examined by the pyrosequencing of rDNA amplicons [48]. Compared to raw milk, in which mesophilic LAB species were the majority, reads of *Streptococcus thermophilus* accounted for about 99% of the sequences found. The 4% and 0.1% of *Vibrio* spp. reads returned by the raw and treated milk, respectively, was not explained [48]. Vibrionaceae read sequences have recently been associated with the new species *Vibrio casei* [103], which was originally isolated from a French, washed-rind cheese [104]. Small numbers of DNA sequences assigned to species never before isolated from milk or dairy products were also noticed, including *Geobacillus toebii* and *Methylobacterium populi*; these may have been environmental contaminants.

3.2.4. Composition and succession of meat-associated microbiota

Community profiling via rDNA pyrosequencing has been used in a meat system to evaluate microbial diversity shifts during storage under different conditions [105]. The microbial data correlated well with the volatile and non-volatile microbial metabolites produced during storage and detected by gas chromatography (SPME-GC-MS) and proton nuclear magnetic resonance (¹H NMR). As expected, the meat microbiota was significantly affected by the storage conditions. The growth of Enterobacteriaceae and *Pseudomonas* spp. was inhibited in modified atmosphere conditions, at least during the first three weeks of storage. Vacuum packaging in nisin-containing bags reduced the growth of *Brochothrix thermosphacta*, a major meat-spoiling organism. Pyrosequencing further revealed complex shifts in most microbial populations; this might have a negative influence on meat quality.

3.3. Spatial and temporal distribution of microbial populations

Bokulich et al. [106] used the Solexa platform technology to study the bacteria and fungi colonizing the surfaces of a pilot plant winery and its manufacturing tools before, during and after harvest over two seasons (Table 2). Under normal conditions, the winery surfaces harbored seasonally fluctuating populations of both types of microorganism. Community composition was also dependent on the production context of each sampling site, and was shaped by the technological practices used and the processing stage. Interestingly, *S. cerevisiae* and other beneficial fermentation-related yeast were detected on the surfaces prior to harvest, while spoilage-related microbes were undetected or detected only at low levels.

In a similar study involving the same technology, the manufacturing and ripening areas of two artisanal cheese factories were analyzed for bacterial and fungal populations [103]. Milk fermentation-associated microbes dominated most of the surfaces, among which *Debaryomyces* and *Lactococcus* were the dominant eukaryotic and prokaryotic organisms respectively. The establishment of these organisms on processing surfaces may play a pivotal role in their transfer between sequential fermentations. Interestingly, similar microbial communities occupied the same surface types, again suggesting the strong selection of microbes on the basis of production technology and stage [103]. Facility-specific microorganisms were also detected, which might influence the sensorial properties of the cheeses they produce.

3.4. Shotgun metagenomics

Analysis of total DNA allows the simultaneous monitoring of bacterial populations and their metabolic potential. In addition, whole DNA sequencing enables a vast array of microbial genes to be identified and annotated, including novel genes and operons.

Pyrosequencing of total microbial DNA from the milk of healthy cows, and that from cows with clinical and subclinical mastitis, has recently been undertaken [68, 107] (Table 2), and significant differences in their microbial compositions reported. Kuen et al. [68] associated healthy samples with larger proportions of *Pseudomonas*, *Psychrobacter*, and *Ralstonia* reads, while in the mastitis samples reads of *Brevundimonas*, *Burkholderia*, *Sphingomonas* and *Stenotrophomonas* were found in greater abundance. Based on these results the authors proposed that some cases of mastitis may be produced by a mixture of microbes rather than a single pathogen. Similar results (although involving different microbial species) were obtained by Bhatt et al. [107]. In the latter study, large numbers of viral sequences, together with sequences of the hosts species *S. aureus*, *E. coli* and *Enterobacter* and *Yersinia* species, were found. The authors speculated that phages might be involved in providing the milk ‘natural resistance’ against pathogens [107].

The first NGS-based metagenomic study of total DNA from a traditional fermented food was performed on kimchi [90] (Table 2). Assignment of metagenomic sequences to functional gene categories revealed that kimchi fermentation was mostly accomplished by heterofermentative LAB species, as reported above in targeted metagenomic studies [89]. Analysis of rDNA sequences among the pyrosequencing reads showed that *Leuconostoc*, *Lactobacillus*, *Weissella* and *Enterococcus* were dominant throughout fermentation. Surprisingly, when the metagenomic reads were mapped onto the database of complete genomes, the two most often recruited genomes

were those of *Leuconostoc mesenteroides* and *Lb. sakei*. However, the metagenome sequences poorly matched the genomes of *Leuconostoc kimchi*, a species isolated from this product, and *Weissella paramesenteroides*, a species that the analysis of rDNA libraries had suggested was dominant throughout fermentation [31]. Similar results were obtained by assembling the pyrosequencing reads into contigs and comparing them against the GenBank non-redundant nucleotide database. Most contigs shared a nucleotide similarity of >97% to the *Leuc. mesenteroides* and *Lb. sakei* genomes. It is noteworthy that a large number of kimchi metagenomic contigs were classified as belonging to bacteriophage DNA, suggesting that a high proportion of bacterial cells either contained integrated prophages or were infected with them [108]. Viral metagenomics could help us understand the ecological role of viruses in the bacterial dynamics of food environments [108, 109].

Pyrosequencing has also been used to compare the composition and diversity of microbial communities in marinated and unmarinated broiler fillets [110]. Bacterial diversity and succession was analyzed by the sequencing of rDNA amplicons and then later by shotgun metagenomics. Smaller populations of *Carnobacterium*, *Vagococcus*, *Brochothrix*, *Clostridium*, Enterobacteriaceae and *Vibrio* were found in marinated meat compared to the unmarinated fillets [110]. *Vagococcus* and *Vibrio*, the predominant communities in the unmarinated samples, had never before been associated with the shelf-life of meat. In addition, increased proportions of LAB species belonging to the Lactobacillaceae and Leuconostocaceae families were encountered in the marinated samples. According to the functional analysis of the metagenomes, over-represented genes in the unmarinated strips were shown to be specific to Gram-negative bacteria. The differences in the genes between the marinated and unmarinated samples were

thought to be due to the different phylogenetic composition of the bacterial communities. The authors concluded that marinating extended both the self-life and sensorial properties of broiler meat by delaying the growth of spoilage microorganisms.

A further shotgun metagenomic DNA study involving pyrosequencing was conducted to identify the prokaryotic and eukaryotic microorganisms involved in cocoa bean fermentation [111]. To determine the best computational approach for estimating the microbial diversity, taxonomic profiling was performed using both similarity-based and composition-based methods involving a vast array of taxonomic profiling tools. Overall, greater community diversity was detected by pyrosequencing than by culturing and previously-used culture-independent methods. The members of the family Lactobacillaceae were found to be the most abundant, with *Lactobacillus* the dominant genus. Subdominant populations of γ -Proteobacteria belonging to the genus *Escherichia* were also recorded. As in other food fermentations, a moderate proportion (0.25% of the total) of viral reads was encountered. Most phage sequences were assigned to those of *Lactobacillus* and enterobacteria, supporting the idea that an interaction exists between bacterial hosts and viral communities [111]. Metagenomic analysis indicated the most abundant yeast to be *Hanseniaspora uvarum*, followed by *Hanseniaspora opuntiae* and *S. cerevisiae* - species that have been commonly associated with cocoa bean fermentation. Other fungal species were also detected, including some which had never before been associated with cocoa fermentation.

3.5. Sequencing of total transcribed RNA: metatranscriptomics

Metatranscriptomics stands for the sequencing and analysis of the total RNA transcripts in an ecosystem under given environmental conditions. The short half-life of

mRNAs - typically a few seconds to a few minutes – actually allows a better view of the microbial activities, metabolic processes and trophic chains occurring within the food at any particular time during manufacture. Metatranscriptomic approaches have been used successfully to examine complex microbial ecosystems such as the rumen [112] and soil [113], and have been suggested promising for studying food fermentations [33]. As yet, however, no food studies using this technique have been reported.

4. The genomics of food-borne microorganisms

NGS has not only contributed towards the study of food microbial ecology but also to the genome sequencing of food-borne microorganisms. The Genome Online Database (GOLD; <http://www.genomesonline.org>), which harbors the most updated genome information, summarizes (at the time of writing, November, 2013) the complete sequence of 2649 bacterial and 151 eukaryotic genomes. These include starter, pathogenic and spoilage strains of hundreds of different species. Analyzing a number of strains of the same species is critical when studying genetic diversity and phenotypic heterogeneity. Genetic analyses provide the sum of all the genes present in a given species (pangenome) and the number of genes shared by all members of that species (core genome). This knowledge is important in understanding the basic biology of pathogens and how to combat them [17], and in developing applications for potentially technologically important microorganisms [114]. For instance, the core genomes of *E. coli* and *Salmonella enterica* strains have been shown to account for 1700 and 2800 genes respectively, while the pangenomes of these species covers 16,000 and 10,000 genes respectively [115]. In contrast, the genome analysis of 14 *Oenococcus oeni*

strains suggested its core genome to be composed of around 1600 genes and its pangenome to contain approximately 1700 [116].

Core genome and pangenome analysis and comparison are powerful tools that have revealed hitherto unknown processes such as aerobic respiration in *Lc. lactis* [117] as well as the presence of pilin-like structures in probiotic lactobacilli [118]. Connections between phenotype, genotype and environmental adaptations can also be predicted from core genome analysis. Knowledge of the core genome could also assist in optimizing culture conditions [119] and/or the production of microorganisms for their use [117]. In contrast, pangenome analysis can be of much help in the search for desirable functional traits. Moreover, genome sequencing can be used to check the safety of microorganisms used as starters or probiotics in the food chain by ruling out the presence of undesirable genes, e.g., those coding for virulence factors and antibiotic resistance.

5. Combination of “omic” techniques

In a recent study, genomic and metagenomic techniques were used to characterize a complex, undefined cheese starter culture for Gouda cheese with a long history of industrial use [109]. Analysis of rDNA sequences revealed the dominance of *Lc. lactis* (99%) and a minor population of *Leuc. mesenteroides* (1%). Culturing and typing techniques identified five genetic lineages of *Lc. lactis* subsp. *cremoris*, two of *Lc. lactis* subsp. *lactis* biovar *diacetylactis*, and one of *Leuc. mesenteroides*. Genome sequencing of all eight lineages allowed the contribution of individual strains to metagenomic data to be tracked during starter propagation and cheese manufacture. Lineage contribution was found to be in good agreement with the results obtained by culturing. Genome analysis was further used to reconstruct *in silico* metabolic maps, with which the

authors evaluated complementary metabolic reactions that might contribute towards species and lineage maintenance. The superimposition of metabolic maps suggested that γ -aminobutyric acid excreted by *Lc. lactis* during the acid stress response might serve as a substrate for succinate formation by *Leuc. mesenteroides* [109]. The presence of 1.15% phage-related DNA sequences in the metagenomic data was also reported. Using strains of *Lc. lactis* as indicators, three lytic phages, two of the p335 type and one of the p936 type, were isolated from the starter mixture. Surprisingly, the *Lc. lactis* strains showed wide variation in phage sensitivity within and between lineages. The stability and dynamics of the microbial community during backslipping propagation of this starter, and throughout cheese manufacture, strongly suggest phage resistance/susceptibility to be a key factor in the prevention of the loss of lineages as environmental conditions change. As in other ecosystems [120], phage predation was proposed to ensure diversity by suppressing abundant strains (“kill the winner” theory), and thus stabilizing overall community functionality [109].

6. Conclusions and future prospects

As in other fields of science, NGS techniques will ultimately revolutionize food microbiology. Pathogen detection, microbial profiling, genetic data mining, genotype-phenotype linking, determining the fate of starters and pathogens over food manufacturing and ripening, and predicting product shelf life etc., will be the realm of these high-throughput technologies. In fact, the microbiology of most foods and food fermentations will shortly be revisited with the aid of one or more of these new techniques. To date, much work has been devoted to testing these techniques in a variety of food and food fermentation ecosystems. They have been mostly used for

descriptive purposes, e.g., in the discovery of novel, low abundance, taxonomic lineages. New genes and operons have already been associated with certain foods and food processes, and a vast panoply of new species has already been associated with some foods. Such knowledge will allow for the determination of critical microbial variables and therefore help better control food quality and safety. In addition, the isolation of strains belonging to newly discovered taxa will ultimately allow their functional characterization, and perhaps their future use as improved starters. Further, the use of NGS in the testing of scientific hypotheses has now begun.

However, current NGS technologies suffer limitations that must be overcome. Much progress needs to be made before reliable data of biological significance can be generated. Recent discoveries and ensuing work have revealed limitations and biases that were previously ignored. Refining the existing technologies should allow the analysis of longer reads, thus permitting genetic identification at lower taxonomic level (which has more biological meaning). Increasing the length of reads to cover the whole of the 16S rRNA gene would undoubtedly improve identification of microbes. Indeed, identification at the species level continues to be crucial in order to obtain data with biological significance. Thus, both sequence length and sequencing accuracy become critical. Enlarged and well-kept databases would also help in accurately determining the phylogenetic position of sequence data. Sequencing technologies currently under development will further increase phylogenetic accuracy. Metagenomic and metatranscriptomic approaches could then be used with confidence to assign sequences to specific species and/or strains and to track their growth in foods. Presumably sequencing cost reductions in the near future will fuel the development of shotgun projects of targeted sequencing, because beyond the phylogenetic data additional

information on genes and gene functionality of the components of the microbial communities can be obtained. New software programs linking phylogeny with function should further enable the direct assessment of community functionality, in an effort to relate microorganisms to the sensorial properties of foods

Acknowledgements

All authors contribute to review the literature, writing the successive drafts, and revise and approve the final version of the submitted manuscript.

Conflict of Interest

The authors declare no conflicts of interest.

Financial support for working at the authors' laboratories was provided by projects from the Spanish Ministry of Economy and Competitiveness (Ref. AGL2011-24300 and RM2011-00005-00-00) to B.M., and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) to C.T.C.C.R., A.M.O.L., and R.S.P.

References

- [1] Smit, G.; Smit, B.A.; Engels, W.J.M. Flavour formation by LAB and biochemical flavour profiling of cheese products. *FEMS Microbiol. Rev.*, **2005**, 29, 591-610.
- [2] Quigley, L.; O'Sullivan, O.; Beresford, T.P.; Ross, R.P.; Fitzgerald, G.F.; Cotter, P.D. Molecular approaches to analyzing the microbial composition of raw milk and raw milk cheeses. *Int. J. Food Microbiol.*, **2011**, 150, 81-94.

- [3] Amann, R.I.; Ludwig, W.; Schleifer, K.-F. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.*, **1995**, *59*, 143-169.
- [4] Ercolini, D.; Moschetti, G.; Blaiotta, G.; Coppola, S. The potential of a polyphasic PCR-DGGE approach in evaluating microbial diversity of natural whey cultures for water-buffalo Mozzarella cheese production: bias of culture-dependent and culture-independent analyzes. *Syst. Appl. Microbiol.*, **2001**, *24*, 610-617.
- [5] Giraffa, G.; Neviani, E. DNA-based, culture-independent strategies for evaluating microbial communities in food-associated ecosystems. *Int. J. Food Microbiol.*, **2001**, *67*, 19-34.
- [6] Jany, J.L.; Barbier, G. Culture-independent methods for identifying microbial communities in cheese. *Food Microbiol.*, **2008**, *25*, 839-848.
- [7] Ampe, F.; ben Omar, N.; Moizan, C.; Wachter, C.; Guyot, J.P. Polyphasic study of the spatial distribution of microorganisms in Mexican pozol, a fermented maize dough, demonstrates the need for cultivation-independent methods to investigate traditional fermentations. *Appl. Environ. Microbiol.*, **1999**, *65*, 5464-5473.
- [8] Cocolin, L.; Manzano, M.; Cantoni, C.; Comi, G. Denaturing gradient gel electrophoresis analysis of the 16S rRNA gene V1 region to monitor dynamic changes in the bacterial population during fermentation of Italian sausages. *Appl. Environ. Microbiol.*, **2001**, *67*, 5113-5121.
- [9] Randazzo, C.L.; Torriani, S.; Akkermans, A.L.D.; de Vos, W.M.; Vaughan, E.E. Diversity, dynamics, and activity of bacterial communities during production of an artisanal Sicilian cheese as evaluated by 16S rRNA analysis. *Appl. Environ. Microbiol.*, **2002**, *68*, 1882-1892.

- [10] Cocolin, L.; Innocente, N.; Biasutti, M.; Comi, G. The late blowing in cheese: a new molecular approach based on PCR and DGGE to study the microbial ecology of the alteration process. *Int. J. Food Microbiol.*, **2004**, 90, 83-91.
- [11] Ercolini, D.; Mauriello, G.; Blaiotta, G.; Moschetti, G.; Coppola, S. PCR-DGGE fingerprints of microbial succession during a manufacture of traditional water buffalo mozzarella cheese. *J. Appl. Microbiol.*, **2004**, 96, 263-270.
- [12] Delbès, C.; Ali-Mandjee, L.; Montel, M.C. 2007. Monitoring bacterial communities in raw milk and cheese by culture-dependent and -independent 16S rRNA gene-based analysis. *Appl. Environ. Microbiol.*, **2007**, 73, 1882-1891.
- [13] Alegría, A.; Álvarez-Martín, P.; Sacristán, N.; Fernández, E.; Delgado, S.; Mayo, B. Diversity and evolution of majority microbial populations during manufacturing and ripening of Casín, a Spanish traditional, starter-free cheese made of raw cow's milk. *International Journal of Food Microbiol.*, **2009**, 136, 44-51.
- [14] Carraro, L.; Maifreni, M.; Bartolomeoli, I.; Martino, M.E.; Novelli, E.; Frigo, F.; Marino, M.; Cardazzo, B. Comparison of culture-dependent and -independent methods for bacterial community monitoring during Montasio cheese manufacturing. *Res. Microbiol.*, **2011**, 162, 231-239.
- [15] Wilmes, P.; Simmons, S.L.; Deneff, V.J.; Banfield, J.F. The dynamic genetic repertoire of microbial communities. *FEMS Microbiol. Rev.*, **2009**, 33, 109-132.
- [16] Solieri, L.; Dakal, T.C.; Giudici, P. Next-generation sequencing and its potential impact on food microbial genomics. *Ann. Microbiol.*, **2012**, 63, 21-37.
- [17] Liu, G.E. Recent applications of DNA sequencing technologies in food, nutrition and agriculture. *Recent Patents Food Nutr. Agricult.*, **2011**, 3, 187-191.

786 [18] Bokulich, N.A.; Mills, D.A. Next-generation approaches to the microbial ecology
787 of food fermentations. *BMB Reports*, **2012**, 45, 377-389

788 [19] Ercolini, D. High-throughput sequencing and metagenomics: moving forward in
789 the culture-independent analysis of food microbial ecology. *Appl. Environ.*
790 *Microbiol.*, **2013**, 79, 3148-3155.

791 [20] Glenn, T.C. Field Guide to Next Generation DNA Sequencers. *Mol. Ecol.*
792 *Resources*, **2011**, 11, 759-769.

793 [21] Liu, L.; Li, Y.; Li, S.; Hu, N.; He, Y.; Pong, R.; Lin, D.; Lu, L.; Law, M.
794 Comparison of Next-Generation Sequencing Systems. *J. Biomed. Biotechnol.*, **2012**,
795 251364.

796 [22] Quail, M.A.; Smith, M.; Coupland, P.; Otto, T.D.; Harris, S.R.; Connor, T.R.;
797 Bertoni, A.; Swerdlow, H.P.; Gu, Y. A tale of three next generation sequencing
798 platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq
799 sequencers. *BMC Genomics*, **2012**, 13, 341.

800 [23] Schadt, E.E.; Turner, S.; Kasarskis, A. A window into third-generation sequencing.
801 *Human Mol. Gen.*, **2010**, 19, R227-240.

802 [24] Margulies, M.; Egholm, M.; Altman, W.E.; Attiya, S.; Bader, J.S.; Bemben, L.A.;
803 Berka, J.; Braverman, M.S.; Chen, Y.J.; Chen, Z.; Dewell, S.B.; Du, L.; Fierro, J.M.;
804 Gomes, X.V.; Godwin, B.C.; He, W.; Helgesen, S.; Ho, C.H.; Irzyk, G.P.; Jando,
805 S.C.; Alenquer, M.L.; Jarvie, T.P.; Jirage, K.B.; Kim, J.B.; Knight, J.R.; Lanza, J.R.;
806 Leamon, J.H.; Lefkowitz, S.M.; Lei, M.; Li, J.; Lohman, K.L.; Lu, H.; Makhijani,
807 V.B.; McDade, K.E.; McKenna, M.P.; Myers, E.W.; Nickerson, E.; Nobile, J.R.;
808 Plant, R.; Puc, B.P.; Ronan, M.T.; Roth, G.T.; Sarkis, G.J.; Simons, J.F.; Simpson,
809 J.W.; Srinivasan, M.; Tartaro, K.R.; Tomasz, A.; Vogt, K.A.; Volkmer, G.A.; Wang,

810 S.H.; Wang, Y.; Weiner, M.P.; Yu, P.; Begley, R.F.; Rothberg, J.M. Genome
811 sequencing in microfabricated high-density picolitre reactors. *Nature*, **2005**, 437,
812 376-380.

813 [25] Dressman, D.; Yan, H.; Traverso, G.; Kinzler, K.W.; Vogelstein, B. Transforming
814 single DNAmolecules into fluorescent magnetic particles for detection and
815 enumeration of genetic variations. *Proc. Ntl. Acad. Sci. USA*, **2003**, 100, 8817–8822.

816 [26] Bentley, D.R.; Balasubramanian, S.; Swerdlow, H.P.; Smith, G.P.; Milton, J.;
817 Brown, C.G.; Hall, K.P.; Evers, D.J.; Barnes, C.L.; Bignell, H.R.; Boutell, J.M.;
818 Bryant, J.; Carter, R.J.; Keira Cheetham, R.; Cox, A.J.; Ellis, D.J.; Flatbush, M.R.;
819 Gormley, N.A.; Humphray, S.J.; Irving, L.J.; Karbelashvili, M.S.; Kirk, S.M.; Li, H.;
820 Liu, X.; Maisinger, K.S.; Murray, L.J.; Obradovic, B.; Ost, T.; Parkinson, M.L.;
821 Pratt, M.R.; Rasolonjatovo, I.M.; Reed, M.T.; Rigatti, R.; Rodighiero, C.; Ross,
822 M.T.; Sabot, A.; Sankar, S.V.; Scally, A.; Schroth, G.P.; Smith, M.E.; Smith, V.P.;
823 Spiridou, A.; Torrance, P.E.; Tzonev, S.S.; Vermaas, E.H.; Walter, K.; Wu, X.;
824 Zhang, L.; Alam, M.D.; Anastasi, C.; Aniebo, I.C.; Bailey, D.M.; Bancarz, I.R.;
825 Banerjee, S.; Barbour, S.G.; Baybayan, P.A.; Benoit, V.A.; Benson, K.F.; Bevis, C.;
826 Black, P.J.; Boodhun, A.; Brennan, J.S.; Bridgham, J. A.; Brown, R.C.; Brown,
827 A.A.; Buermann, D.H.; Bundu, A.A.; Burrows, J.C.; Carter, N.P.; Castillo, N.;
828 Chiara, E.C.M.; Chang, S.; Neil Cooley, R.; Crake, N.R.; Dada, O.O.; Diakoumakos,
829 K.D.; Dominguez-Fernandez, B.; Earnshaw, D.J.; Egbujor, U.C.; Elmore, D.W.;
830 Etchin, S.S.; Ewan, M.R.; Fedurco, M.; Fraser, L.J.; Fuentes Fajardo, K.V.; Scott
831 Furey, W.; George, D.; Gietzen, K.J.; Goddard, C.P.; Golda, G.S.; Granieri, P.A.;
832 Green, D.E.; Gustafson, D.L.; Hansen, N.F.; Harnish, K.; Haudenschild, C.D.;
833 Heyer, N.I.; Hims, M.M.; Ho, J.T.; Horgan, A.M.; Hoschler, K.; Hurwitz, S.; Ivanov,

834 D.V.; Johnson, M.Q.; James, T.; Huw Jones, T.A.; Kang, G.D.; Kerelska, T.H.;
 835 Kersey, A.D.; Khrebtukova, I.; Kindwall, A.P.; Kingsbury, Z.; Kokko-Gonzales, P.I.;
 836 Kumar, A.; Laurent, M.A.; Lawley, C.T.; Lee, S.E.; Lee, X.; Liao, A.K.; Loch, J.A.;
 837 Lok, M.; Luo, S.; Mammen, R.M.; Martin, J.W.; McCauley, P.G.; McNitt, P.; Mehta,
 838 P.; Moon, K.W.; Mullens, J.W.; Newington, T.; Ning, Z.; Ling Ng, B.; Novo, S.M.;
 839 O'Neill, M.J.; Osborne, M.A.; Osnowski, A.; Ostadan, O.; Paraschos, L.L.;
 840 Pickering, L.; Pike, A.C.; Chris Pinkard, D.; Pliskin, D.P.; Podhasky, J.; Quijano,
 841 V.J.; Raczy, C.; Rae, V.H.; Rawlings, S.R.; Chiva Rodriguez, A.; Roe, P.M.; Rogers,
 842 J.; Rogert Bacigalupo, M.C.; Romanov, N.; Romieu, A.; Roth, R.K.; Rourke, N.J.;
 843 Ruediger, S.T.; Rusman, E.; Sanches-Kuiper, R.M.; Schenker, M.R.; Seoane, J.M.;
 844 Shaw, R.J.; Shiver, M.K.; Short, S.W.; Sizto, N.L.; Sluis, J.P.; Smith, M.A.; Ernest
 845 Sohna Sohna, J.; Spence, E.J.; Stevens, K.; Sutton, N.; Szajkowski, L.; Tregidgo,
 846 C.L.; Turcatti, G.; Vandevondele, S.; Verhovsky, Y.; Virk, S.M.; Wakelin, S.;
 847 Walcott, G.C.; Wang, J.; Worsley, G.J.; Yan, J.; Yau, L.; Zuerlein, M.; Mullikin,
 848 J.C.; Hurles, M.E.; McCooke, N.J.; West, J.S.; Oaks, F.L.; Lundberg, P.L.;
 849 Klenerman, D.; Durbin, R.; Smith, A.J. Accurate whole human genome sequencing
 850 using reversible terminator chemistry. *Nature*, **2008**, 456, 53-9.
 851 [27] Valouev, A.; Ichikawa, J.; Tonthat, T.; Stuart, J.; Ranade, S.; Peckham, H.; Zeng,
 852 K.; Malek, J.A.; Costa, G.; MaKernan, K.; Sidow, A.; Fire, A.; Johnson, S.M. A
 853 high-resolution, nucleosome position map of *c. elegans* reveals a lack of universal
 854 sequence-dictated positioning. *Genome Res.*, **2008**, 18, 1051-1063.
 855 [28] Rothberg, J.M.; Hinz, W.; Rearick, T.M.; Schultz, J.; Mileski, W.; Davey, M.;
 856 Leamon, J.H.; Johnson, K.; Milgrew, M.J.; Edwards, M.; Hoon, J.; Simons, J.F.;
 857 Marran, D.; Myers, J.W.; Davidson, J.F.; Branting, A.; Nobile, J.R.; Puc, B.P.; Light,

D.; Clark, T.A.; Huber, M.; Branciforte, J.T.; Stoner, I.B.; Cawley, S.E.; Lyons, M.;
 Fu, Y.; Homer, N.; Sedova, M.; Miao, X.; Reed, B.; Sabina, J.; Feierstein, E.;
 Schorn, M.; Alanjary, M.; Dimalanta, E.; Dressman, D.; Kasinskas, R.; Sokolsky, T.;
 Fidanza, J.A.; Namsaraev, E.; McKernan, K.J.; Williams, A.; Roth, G.T.; Bustillo, J.
 An integrated semiconductor device enabling non-optical genome sequencing.
 Nature, **2011**, 475:348-352.

[29] Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.;
 Baybayan, P.; Bettman, B.; Bibillo, A.; Bjornson, K.; Chaudhuri, B.; Christians, F.;
 Cicero, R.; Clark, S.; Dalal, R.; Dewinter, A.; Dixon, J.; Foquet, M.; Gaertner, A.;
 Hardenbol, P.; Heiner, C.; Hester, K.; Holden, D.; Kearns, G.; Kong, X.; Kuse, R.;
 Lacroix, Y.; Lin, S.; Lundquist, P.; Ma, C.; Marks, P.; Maxham, M.; Murphy, D.;
 Park, I.; Pham, T.; Phillips, M.; Roy, J.; Sebra, R.; Shen, G.; Sorenson, J.; Tomaney,
 A.; Travers, K.; Trulson, M.; Vieceli, J.; Wegener, J.; Wu, D.; Yang, A.; Zaccarin,
 D.; Zhao, P.; Zhong, F.; Korlach, J.; Turner, S. Real-time DNA sequencing from
 single polymerase molecules. Science, **2009**, 323, 133-138.

[30] Escalante, A.; Wachter, C.; Farrés, A. Lactic acid bacterial diversity in the
 traditional Mexican fermented dough pozol as determined by 16S rDNA sequence
 analysis. Int. J. Food Microbiol., **2001**, 64, 21-31.

[31] Kim, M.; Chun, J. Bacterial community structure in kimchi, a Korean fermented
 vegetable food, as revealed by 16S rRNA gene analysis. Int. J. Food Microbiol.,
2005, 103, 91-96.

[32] Ninane, V.; Mukandayambaje, R.; Berben, G. Identification of lactic acid bacteria
 within the consortium of a kefir grain by sequencing 16S rDNA variable regions. J.
 AOAC Int., **2007**, 90, 1111-1117.

- [33] van Hijum, S.A.F.T.; Vaughan, E.E.; Vogel, R.F. Application of state-of-art sequencing technologies to indigenous food fermentations. *Curr. Opin. Biotechnol.*, **2013**, 24, 178-186.
- [34] Linnarsson, S. Recent advances in DNA sequencing methods - general principles of sample preparation. *Exp. Cell Res.*, **2010**, 316, 1339-1343.
- [35] Bonaiti, C.; Parayre, S.; Irlinger, F. Novel extraction strategy of ribosomal RNA and genomic DNA from cheese for PCR-based investigations. *Int. J. Food Microbiol.*, **2006**, 107, 171-179.
- [36] Quail, M.A.; Kozarewa, I.; Smith, F.; Scally, A.; Stephens, P.J.; Durbin, R.; Swerdlow, H.; Turner, D.J. A large genome center's improvements to the Illumina sequencing system. *Nat. Methods*, **2008**, 5, 1005-1010.
- [37] Masoud, W.; Vogensen, F.K.; Lillevang, S.; Al-Soud, W.A.; Sørensen, S.J.; Jakobsen, M. The fate of indigenous microbiota, starter cultures, *Escherichia coli*, *Listeria innocua* and *Staphylococcus aureus* in Danish raw milk and cheeses determined by pyrosequencing and quantitative real-time (qRT)-PCR. *Int. J. Food Microbiol.*, **2012**, 153, 192-202.
- [38] Cocolin, L.; Alessandria, V.; Botta, C.; Gorra, R.; de Filippis, F.; Ercolini, D.; Rantsiou, K. NaOH-Debittering induces changes in bacterial ecology during table olives fermentation. *PLoS One*, **2013**, 8, e69074.
- [39] Schmieder, R.; Lim, Y.W.; Edwards, R. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics*, **2012**, 28, 433-435.
- [40] Blazewicz, S.J.; Barnard, R.L.; Daly, R.A.; Firestone, M.K. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME J.*, **2013**, 7, 2061-2068.

- [41] Mader, U.; Nicolas, P.; Richard, H.; Bessieres, P.; Aymerich, S. Comprehensive identification and quantification of microbial transcriptomes by genome-wide unbiased methods. *Curr. Opin. Biotechnol.*, **2011**, 22, 32-41.
- [42] Simon, C.; Daniel, R. Metagenomic analyses: past and future trends. *Appl. Environ. Microbiol.*, **2011**, 77, 1153-1161.
- [43] Nocker, A.; Cheung, C.-Y.; Camper, A.K. Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells. *J. Microbiol. Methods*, **2006**, 67, 310-320.
- [44] Soergel, D.A.; Dey, N.; Knight, R.; Brenner, S.E. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J.*, **2012**, 6, 1440-1444.
- [45] Roh, S.W.; Kim, K.-H.; Nam, Y.-D.; Chang, H.-W.; Park, E.-J.; Bae, J.-W. Investigation of archaeal and bacterial diversity in fermented seafood using barcoded pyrosequencing. *The ISME J.*, **2010**, 4, 1-16.
- [46] Bokulich, N.A.; Joseph, C.M.; Allen, G.; Benson, A.K.; Mills, D.A. Next-generation sequencing reveals significant bacterial diversity of botrytized wine. *PLoS One*, **2012**, 7, e36357.
- [47] Leite, A.M.; Mayo, B.; Rachid, C.T.; Peixoto, R.S.; Silva, J.T.; Paschoalin, V.M.; Delgado, S. Assessment of the microbial diversity of Brazilian kefir grains by PCR-DGGE and pyrosequencing analysis. *Food Microbiol.*, **2012**, 31, 215-221.
- [48] Delgado, S.; Rachid, C.T.C.C.; Fernández, E.; Rychlik, T.; Alegría, A.; Peixoto, R.S.; Mayo, B. Diversity of thermophilic bacteria in raw, pasteurized and selectively-cultured milk, as assessed by culturing, PCR-DGGE and pyrosequencing. *Food Microbiol.*, **2013**, 36, 103-111.

- [49] Ercolini, D.; Pontonino, E.; de Filippis, F.; Minervani, F.; La Stora, A.; Gobbetti, M.; Di Cagno, R. Microbial ecology dynamics during rye and wheat sourdough preparation. *Appl. Environ. Microbiol.*, **2013**, 79, 7827-7836.
- [50] Lee, S.H.; Jung, J.Y.; Jeon, C.O. Effects of temperature on microbial succession and metabolite change during saeu-jeot fermentation. *Food Microbiol.*, **2014**, 38, 16-25.
- [51] Liu, Z.; Lozupone, C.; Hamady, M.; Bushman, F.D.; Knight, R. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res.*, **2007**, 35, e120.
- [52] Petrosino, J.F.; Highlander, S.; Luna, R.A.; Gibbs, R.A.; Versalovic, J. Metagenomic pyrosequencing and microbial identification. *Clin. Chem.*, **2009**, 55, 856-866.
- [53] Rajendhran, J.; Gunasekaran, P. Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. *Microbiol Res.*, **2011**, 166, 99-110.
- [54] Ansorge, W.J. Next-generation DNA sequencing techniques. *Nat. Biotechnol.*, **2009**, 25, 195-203.
- [55] Renouf, V.; Claisse, O.; Miot-Sertier, C.; Lonvaud-Funel, A. Lactic acid bacteria evolution during winemaking: use of *rpoB* gene as a target for PCR-DGGE analysis. *Food Microbiol.*, **2006**, 23: 136-145.
- [56] Volokhov, D.V.; Simonyan, V.; Davidson, M.K.; Chizhikov, V.E. RNA polymerase beta subunit (*rpoB*) gene and the 16S-23S rRNA intergenic transcribed spacer region (ITS) as complementary molecular markers in addition to the 16S rRNA gene for phylogenetic analysis and identification of the species of the family *Mycoplasmataceae*. *Mol. Phylogenet. Evol.*, **2012**, 62, 515-528.

- [57] Sibley, C.D.; Peirano, G.; Church, D.L. Molecular methods for pathogen and microbial community detection and characterization: current and potential application in diagnostic microbiology. *Infect. Genet. Evol.*, **2012**, *12*, 505-521.
- [58] Kembel, S.W.; Wu, M.; Eisen, J.; Green, J.L. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Computational Biol.*, **2012**, *8*, e1002743.
- [59] Větrovský, T.; Baldrian, P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PloS One*, **2013**, *8*, e57923.
- [60] Acinas, S.G.; Sarma-rupavtarm, R.; Klepac-ceraj, V.; Polz, M.F. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *Appl. Environ. Microbiol.*, **2005**, *71*, 8966-8969.
- [61] Anderson, I.C.; Campbell, C.D.; Prosser, J.I. Potential bias of fungal 18S rDNA and internal transcribed spacer polymerase chain reaction primers for estimating fungal biodiversity in soil. *Environ. Microbiol.*, **2003**, *5*, 36-47.
- [62] Bokulich, N.A.; Mills, D.A. Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. *Appl. Environ. Microbiol.*, **2103**, *79*, 2519-2526.
- [63] Groenewald, M.; Robert, V.; Smith, M.T. The value of the D1/D2 and internal transcribed spacers (ITS) domains for the identification of yeast species belonging to the genus *Yamadazyma*. *Persoonia*, **2011**, *26*, 40-46
- [64] Nilsson, R.H.; Ryberg, M.; Abarenkov, K.; Sjökvist, E.; Kristiansson, E. The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. *FEMS Microbiol. Lett.*, **2009**, *296*, 97-101.

978 [65] Kruger, D.; Kapturska, D.; Fischer, C.; Daniel, R.; Wubet, T. Diversity measures in
979 environmental sequences are highly dependent on alignment quality-data from ITS
980 and new LSU primers targeting basidiomycetes. *PLoS One*, **2012**, 7, e32139.

981 [66] Binladen, J.; Gilbert, M.T.; Bollback, J.P.; Panitz, F.; Bendixen, C.; Nielsen, R.;
982 Willerslev, E. The use of coded PCR primers enables high-throughput sequencing of
983 multiple homolog amplification products by 454 parallel sequencing. *PLoS One*,
984 **2007**, 2, e197.

985 [67] Mamanova, L.; Coffey, A.J.; Scott, C.E.; Kozarewa, I.; Turner, E.H.; Kumar, A.;
986 Howard, E.; Shendure, J.; Turner, D.J. Target-enrichment strategies for next-
987 generation sequencing. *Nature Methods*, **2010**, 7, 111-118.

988 [68] Kuehn, J.S.; Gorden, P.J.; Munro, D.; Rong, R.; Dong, Q.; Plummer, P.J.; Wang,
989 C.; Phillips, G.J. Bacterial community profiling of milk samples as a means to
990 understand culture-negative bovine clinical mastitis. *PLoS One*, **2013**, 8, e61959.

991 [69] Silander, K.; Saarela, J. Whole genome amplification with Phi29 DNA polymerase
992 to enable genetic or genomic analysis of samples of low DNA yield. *Methods Mol.*
993 *Biol.*, 2008, 439, 1-18.

994 [70] Sul, W.J.; Cole, J.R.; Jesus, E.D.C.; Wang, Q.; Farris, R.J.; Fish, J.A.; Tiedje, J.M.
995 Bacterial community comparisons by taxonomy-supervised analysis independent of
996 sequence alignment and clustering. *Proc. Natl. Acad. Sci. USA*, **2011**, 108, 14637-
997 14642.

998 [71] Cole, J.R.; Wang, Q.; Cardenas, E.; Fish, J.; Chai, B.; Farris, R.J.; Kulam-Syed-
999 Mohideen, A.S.; McGarrell, D.M.; Marsh, T.; Garrity, G.M.; Tiedje, J.M. The
1000 Ribosomal Database Project: improved alignments and new tools for rRNA analysis.
1001 *Nucleic Acids Res.*, **2009**, 37, D141–D145.

1002 [72] Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Pelies, J.;
1003 Glöckner, F.O. The SILVA ribosomal RNA gene database project: improved data
1004 processing and web-based tools. *Nucleic Acids Res.*, **2013**, 41, D590-D596.

1005 [73] DeSantis, T.Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E.L.; Keller, K.;
1006 Huber, T.; Dalevi, D.; Hu, P.; Andersen, G.L. Greengenes, a chimera-checked 16S
1007 rRNA gene database and workbench compatible with ARB. *Appl. Environ.*
1008 *Microbiol.*, **2006**, 72, 5069-5072.

1009 [74] Aziz, R.; Bartels, D.; Best, A.; DeJongh, M.; Disz, T.; Edwards, R.; Formsma, K.;
1010 Gerdes, S.; Glass, E.; Kubal, M.; Meyer, F.; Olsen, G.; Olson, R.; Osterman, A.;
1011 Overbeek, R.; McNeil, L.; Paarmann, D.; Paczian, T.; Parrello, B.; Pusch, G.; Reich,
1012 C.; Stevens, R.; Vassieva, O.; Vonstein, V.; Wilke, A.; Zagnitko, O. The RAST
1013 Server: Rapid Annotations using Subsystems Technology. *BMC genomics*, **2008**, 9,
1014 75.

1015 [75] Meyer, F.; Paarmann, D.; D'Souza, M.; Olson, R.; Glass, E.M.; Kubal, M.; Paczian,
1016 T.; Rodriguez, A.; Stevens, R.; Wilke, A.; Wilkening, J.; Edwards, R.A. The
1017 metagenomics RAST server-a public resource for the automatic phylogenetic and
1018 functional analysis of metagenomes. *BMC bioinformatics*, **2008**, 9, 386.

1019 [76] Rachid, C.T.C.C.; Santos, A.L.; Piccolo, M.C.; Balieiro, F.C.; Coutinho, H.L.C.;
1020 Peixoto, R.S.; Tiedje, J.M.; Rosado, A.S. Effect of sugarcane burning or green
1021 harvest methods on the Brazilian Cerrado soil bacterial community structure. *PloS*
1022 *One*, **2013**, 8, e59342.

1023 [77] Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.;
1024 Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; Sahl, J.W.; Stres, B.;
1025 Thallinger, G.G.; van Horn, D.J.; Weber, C.F. Introducing mothur: Open source,

platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*, **2009**, 75, 7537-7541.

[78] Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Pena, A.G.; Goodrich, J.K.; Gordon, J.I.; Huttley, G.A.; Kelley, S.T.; Knights, D.; Koenig, J.E.; Ley, R.E.; Lozupone, C.A.; McDonald, D.; Muegge, B.D.; Pirrung, M.; Reeder, J.; Sevinsky, J.R.; Turnbaugh, P.J.; Walters, W.A.; Widmann, J.; Yatsunenko, T.; Zaneveld, J.; Knight, R. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*, **2010**, 7, 335-336.

[79] Quince, C.; Lanzen, A.; Davenport, R.J.; Turnbaugh, P.J. Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, **2011**, 12, 38.

[80] Quince, C.; Lanzén, A.; Curtis, T.P.; Davenport, R.J.; Hall, N.; Head, I.M.; Read, L.F.; Sloan, W.T. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature Methods*, **2009**, 6, 639-641.

[81] Reeder, J.; Knight, R. Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nat. Methods*, 2010, 7, 668-669.

[82] Sogin, M. L., Morrison, H.; Mclellan, S.; Welch, D.M.; Huse, S. The rare biosphere : sorting out fact from fiction. *Genome Biolo.*, **2010**, 11, SI19.

[83] Langille, M.G.; Zaneveld, J.; Caporaso, J.G.; McDonald, D.; Knights, D.; Reyes, J.A.; Clemente, J.C.; Burkepile, D.E.; Vega Thurber, R.L.; Knight, R.; Beiko, R.G.; and Huttenhower, C. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.*, **2013**, 31, 814-821.

[84] Humblot, C.; and Guyot, J.P. Pyrosequencing of tagged 16S rRNA gene amplicons for rapid deciphering of the microbiomes of fermented foods such as pearl millet slurries. *Appl. Environ. Microbiol.*, **2009**, 75, 4354-4361.

- [85] Kim, Y.S.; Kim, M.C.; Kwon, S.W.; Kim, S.J.; Park, I.C.; Ka, J.O.; Weon, H.Y. Analyses of bacterial communities in meju, a Korean traditional fermented soybean bricks, by cultivation-based and pyrosequencing methods. *J. Microbiol.*, **2011**, 49, 340-348.
- [86] Nam, Y.-D.; Yi, S.-H.; Lim, S.-I. Bacterial diversity of *cheonggukjang*, a traditional Korean fermented food, analyzed by barcoded pyrosequencing. *Food Control*, **2012**, 28, 135-142.
- [87] Nam, Y.-D.; Lee, S.-Y.; Lim, S.-I. Microbial community analysis of Korean soybean pastes by next-generation sequencing. *Int. J. Food Microbiol.*, **2012**, 155, 36-42.
- [88] Nam, Y.-D.; Park, S.-I.; Lim, S.-I. Microbial composition of the Korean traditional food “*kochujang*” analyzed by a massive sequencing technique. *J. Food Sci.*, **2012**, 77, 250-256.
- [89] Park, E.-J.; Chun, J.; Cha, C.-J.; Park, W.-S.; Jeon, C.O.; Bae, J.-W. Bacterial community analysis during fermentation of ten representative kinds of “kimchi” with barcoded pyrosequencing. *Food Microbiol.*, **2012**, 30, 197-204.
- [90] Jung, J.Y.; Lee, S.H.; Kim, J.M.; Park, M.S.; Bae, J.-I.; Hahn, Y.; Madsen, E.L.; Jeon, C.O. Metagenomic analysis of “kimchi”, a traditional Korean fermented food. *Appl. Environ. Microbiol.*, **2011**, 77, 2264-2274.
- [91] Koyanagi, T.; Kiyohara, M.; Matsui, H.; Yamamoto, K.; Kondo, T.; Katayama, T.; Kumagai, H. Pyrosequencing survey of the microbial diversity of 'narezushi', an archetype of modern Japanese sushi. *Lett. Appl. Microbiol.*, **2011**, 53, 635-640.

1072 [92] Sakamoto, N.; Tanaka, S.; Sonomoto, K.; Nakayama, J. 16S rRNA
1073 pyrosequencing-based investigation of the bacterial community in nukadoko, a
1074 pickling bed of fermented rice bran. *Int. J. Food Microbiol.*, **2011**, 144, 352-359.

1075 [93] Koyanagi, T.; Nakagawa, A.; Kiyohara, M.; Matsui, H.; Yamamoto, K.; Barla, F.;
1076 Take, H.; Katsuyama, Y.; Tsuji, A.; Shijimaya, M.; Nakamura, S.; Minami, H.;
1077 Enomoto, T.; Katayama, T.; Kumagai, H. Pyrosequencing analysis of microbiota in
1078 kaburazushi, a traditional medieval sushi in Japan. *Biosci. Biotechnol. Biochem.*,
1079 **2013**, 77, 2125-2130.

1080 [94] Jung, M.J.; Nam, Y.D.; Roh, S.W.; Bae, J.W. Unexpected convergence of fungal
1081 and bacterial communities during fermentation of traditional Korean alcoholic
1082 beverages inoculated with various natural starters. *Food Microbiol.*, **2012**, 30, 112-
1083 123.

1084 [95] Bokulich, N.A.; Bamforth, C.W.; Mills, D.A. Brewhouse-resident microbiota are
1085 responsible for multi-stage fermentation of American coolship ale. *PLoS One*, **2012**,
1086 7, e35507.

1087 [96] Oikonomou, G.; Machado, V.S.; Santisteban, C.; Schukken, Y.H.; Bicalho, R.C.
1088 Microbial diversity of bovine mastitis milk as described by pyrosequencing of
1089 metagenomic 16S rDNA. *PLoS One*, **2012**, 7, e47671.

1090 [97] Dobson, A.; O'sullivan, O.; Cotter, P.D.; Ross, P.; Hill, C. High-throughput
1091 sequence-based analysis of the bacterial composition of kefir and an associated kefir
1092 grain. *FEMS Microbiol. Lett.*, **2011**, 320, 56-62.

1093 [98] Masoud, W.; Takamiya, M.; Vogensen, F.K.; Lillevang, S.; Al-Soud, W.A.;
1094 Sørensen, S.J.; Jakobsen, M. Characterization of bacterial populations in Danish raw

1095 milk cheeses made with different starter cultures by denaturing gradient gel
1096 electrophoresis and pyrosequencing. J. Dairy J., **2010**, 21, 142-148.

1097 [99] Alegría, A.; Szczesny, P.; Mayo, B.; Bardowski, J.; Kowalczyk, M. Biodiversity in
1098 Oscypek, a traditional Polish cheese, determined by culture-dependent and -
1099 independent approaches. Appl. Environ. Microbiol., **2012**, 78, 1890-1898.

1100 [100] Quigley, L.; O’Sullivan, O.; Beresford, T.P.; Ross, R.P.; Fitzgerald, G.F.; Cotter,
1101 P.D. High-throughput sequencing for detection of subpopulations of bacteria not
1102 previously associated with artisanal cheeses. Appl. Environ. Microbiol., **2012**, 78,
1103 5717-5723.

1104 [101] Lusk, T.S.; Ottensen, A.R.; White, J.R.; Allard, M.W.; Brown, E.W.; Kase, J.E.
1105 Characterization of microflora in Latin-style cheeses by next-generation sequencing
1106 technology. BCM Microbiol., **2012**, 12, 254.

1107 [102] Ercolini, D.; De Filippis, F.; La Stora, A.; Iacono, M. “Remake” by high-
1108 throughput sequencing of the microbiota involved in the production of water buffalo
1109 Mozzarella cheese. Appl. Environ. Microbiol. **2012**, 78, 8142-8145.

1110 [103] Bokulich, N.A.; Mills, D.A. Facility-specific “house”microbiome drives
1111 microbial landscapes of artisan cheesemaking plants. Appl. Environ. Microbiol.,
1112 **2013**, 79, 5214-5223.

1113 [104] Blicher, A.; Neuhaus, K.; Scherer, S. *Vibrio casei* sp. nov., isolated from the
1114 surface of two French red smear soft cheeses. Int. J. Syst. Evol. Microbiol., **2010**, 60,
1115 1745-1749.

1116 [105] Ercolini, D.; Ferrocino, I.; Nasi, A.; Ndagijimana, M.; Vernocchi, P.; La Stora,
1117 A.; Laghi, L.; Mauriello, G.; Guerzoni, M.E.; and Villani, F. Monitoring of microbial

metabolites and bacterial diversity in beef stored under different packaging conditions. *Appl. Environ. Microbiol.*, **2011**, 77, 7372-7381.

[106] Bokulich, N.A.; Ohta, M.; Richardson, P.M.; Mills, D.A. Monitoring seasonal changes in winery-resident microbiota. *PLoS One*, **2013**, 8, e66437.

[107] Bhatt, V.D.; Ahir, V.B.; Koringa, P.G.; Jakhesara, S.J.; Rank, D.N.; Nauriyal, D.S.; Kunjadia, A.P.; Joshi, C.G. Milk microbiome signatures of subclinical mastitis-affected cattle analysed by shotgun sequencing. *J. Appl. Microbiol.*, **2012**, 112, 639-650.

[108] Park, E.-J.; Kim, K.-H.; Abell, G.C.J.; Kim, M.-S.; Roh, S.W.; Bae, J.-W. Metagenomic analysis of the viral communities in fermented foods. *Appl. Environ. Microbiol.*, **2011**, 77, 1284-1291.

[109] Erkus, O.; de Jager, V.C.L.; Spus, M.; van Alen-Boeritger, I.J.; van Rijswijk, I.M.H.; Hazelwood, L.; Janssen, P.W.M.; van Hijum, S.A.F.T.; Kleerebezem, M.; Smid, E.J. Multifactorial diversity sustains microbial community stability. *The ISME J.*, **2013**, 7, 2126-2136

[110] Nieminen, T.T.; Koskinen, K.; Laine, P.; Hultman, J.; Säde, E.; Paloranta, A.; Johansson, P.; Björkroth, J.; Auvinen, P. Comparison of microbial communities in marinated and unmarinated broiler meat by metagenomics. *Int. J. Food Microbiol.*, **2012**, 157, 142-149.

[111] Illegheems, K.; De Vuyst, L.; Papalexandratou, Z.; Weckx, S. Phylogenetic analysis of a spontaneous cocoa bean fermentation metagenome reveals new insights into its bacterial and fungal community diversity. *PLoS One*, **2012**, 7, e38040.

[112] Wang, P.; Qi, M.; Barboza, P.; Leigh, M.B.; Ungerfeld, E.; Selinger, L.B.; McAllister, T.A.; Forster, R.J. Isolation of high-quality total RNA from rumen

1142 anaerobic bacteria and fungi, and subsequent detection of glycoside hydrolases. Can.
 1143 J. Microbiol., **2011**, 57, 590-598.

1144 [113] Carvalhais, L.C.; Schenk, P.M. Sample processing and cDNA preparation for
 1145 microbial metatranscriptomics in complex soil communities. Methods Enzymol.,
 1146 **2013**, 531, 251-267.

1147 [114] Garrigues, C.; Johansen, E.; Crittenden, R. Pangenomics-an avenue to improved
 1148 industrial starters and probiotics. Curr. Opin. Biotechnol., **2013**, 24, 187-191.

1149 [115] Jacobsen, A.; Hendriksen, R.S.; Aaresturp, F.M.; Ussery, D.W.; Friis, C. The
 1150 *Salmonella enterica* pan-genome. Microb. Ecol., **2011**, 62, 487-504.

1151 [116] Borneman, A.R.; McCarthy, J.M.; Chambers, P.J.; Bartowsky, E.J. Comparative
 1152 analysis of the *Oenococcus oeni* pan genome reveals genetic diversity in industrially-
 1153 relevant pathways. BMC Genomics, **2012**, 13, 373.

1154 [117] Duwat, P.; Sourice, S.; Cesselin, B.; Lamberet, G.; Vido, K.; Gaudu, P.; Le Loir,
 1155 Y.; Violet, F.; Loubière, P.; Gruss, A. Respiration capacity of the fermenting
 1156 bacterium *Lactococcus lactis* and its positive effects on growth and survival. J.
 1157 Bacteriol., **2001**, 183, 4509-4516.

1158 [118] Kankainen, M.; Paulin, L.; Tynkkynen, S.; von Ossowski, I.; Reunanen, J.;
 1159 Partanen, P.; Satokari, R.; Vesterlund, S.; Hendrickx, A.P.A.; Lebeer, S.; De
 1160 Keersmaecker, S.C.J.; Vanderleyden, J.; Hämäläinen, T.; Laukkanen, S.; Salovuori,
 1161 N.; Ritari, J.; Alatalo, E.; Korpela, R.; Mattila-Sandholm, T.; Lassig, A.; Hatakka,
 1162 K.; Kinnunen, K.T.; Karjalainen, H.; Saxelin, M.; Laakso, K.; Surakka, A.; Palva,
 1163 A.; Salusjärvi, T.; Auvinen, P.; de Vos, W.M. Comparative genomic analysis of
 1164 *Lactobacillus rhamnosus* GG reveals pili containing a human-mucus binding protein.
 1165 Proc. Natl. Acad. Sci. USA, **2009**, 106, 17193–17198.

- 1166 [119] Handelsman, J. Metagenomics: application of genomics to uncultured
1167 microorganisms. *Microbiol. Mol. Biol. Rev.*, **2004**, 68, 669-685.
- 1168 [120] Rodríguez-Valera, F., Martín-Cuadrado, A.B., Rodríguez-Brito, B., Pasic, L.,
1169 Thingstad, T.F., Rohwer, F., Mira, A. Explaining microbial population genomics
1170 through phage predation. *Nat. Rev. Microbiol.*, **2009**, 7, 828-836.

Table 1.- Usual terms utilized in most omic techniques.

| Term | Definition |
|-------------------------|---|
| Barcodes | Short nucleotides sequences merged with primers and/or adaptors allowing simultaneous sequencing of DNA from multiple samples and further separation <i>in silico</i> |
| Denoising | Quality processing applied to 16S rDNA reads, correcting the “noise” (errors) artificially generated during sequencing |
| Sequencing trimming | Processing of sequencing reads, which includes the removal of primers and barcodes, deletion of a given sequence region and elimination of low quality and very short reads |
| Assembly | Partial reconstruction of genes or genomes by aligning and merging short sequencing reads |
| OTU | Operational taxonomic unit. A cluster of sequences within a given similarity cut-off (e.g. 3%, which is usually utilized to define bacterial species by 16S rDNA) |
| Binning | Separation of all fragments originating from a common taxon or OTU |
| Phylogenetic assignment | Assignment of each sequence or OTU to its known closest relative organism |
| Coverage | Means how deep was the sequencing effort in sampling a given community; number of times a nucleotide is read during sequencing |
| Richness estimators | OTU, Ace, Chao1 - Estimate the number of species (or genus, orders etc.) in a given sample by different methods |
| Diversity estimators | e.i. Shannon – Estimate the diversity, taking into account the number of species and how even they are distributed |
| α -diversity | Measures the diversity associated with a single sample (e.g. OTU number, Shannon Index, rarefaction curve, etc.) |
| β -diversity | Measures the diversity among samples (e.g. heatmaps, venn diagrams, similarity trees, PCAs, ordinations etc.) |
| Rarefaction curve | Curve representing the richness of the sample according to the number of sequences. The shape of the curve reflects the sample diversity |

Table 2.- Summary of NGS projects analyzing food-associated microbiotas, including working conditions and major findings.

| Food sample/source | Sequencing project | Amplicon length (bp) | Sequencing platform | Taxonomic resolution | Database used ^a | Major findings | Reference |
|---|---|----------------------|---------------------|----------------------|----------------------------|---|-----------|
| Material of plant origin | | | | | | | |
| Fermentation of African pearl millets | V3 16S rDNA | >180 | 454 FLX | Genus | RDP | <i>Weissella</i> , <i>Pediococcus</i> and <i>Lactobacillus</i> were the only genera after 24 h of fermentation; high intersample variability | [84] |
| Fermented shrimp, kimchi and sauerkraut | V1-V3 16S rDNA, total DNA | >300 na | 454 FLX | Genus | RDP | Viral and hosts communities; discrepancy on phage hosts via homology comparison and rDNA sequencing | [86] |
| Fermentation of “kimchi” | Total DNA | na | 454 FLX | Species | RDP and MG-RAST | <i>Leuconostoc</i> , <i>Lactobacillus</i> and <i>Weissella</i> dominant organisms; <i>Leuc. mesenteroides</i> and <i>Lb. sakei</i> genomes highly represented; phage-related sequences | [90] |
| Fermented sushi (“narezushi”) | V1-V2 16S rDNA | >300 | 454 FLX | Genus | RDP | Species of <i>Lactobacillus</i> and <i>Pediococcus</i> always present; occasional presence of genera of LAB and other bacterial groups | [91] |
| Fermented rice brand mash (“nukadoko”) | V6-V8 16S rDNA | >400 | 454 FLX | Genus/species | RDP | Rice-associated bacteria are replaced by <i>Lb. namurensis</i> , <i>Lb. acetotolerans</i> , and some other <i>Lactobacillus</i> during fermentation | [92] |
| Wine made from botrytized grapes | Separate V4 and V5 16S rDNA | >150 | Illumina | Family/genus | RDP | Similar community structure with V4 and V5 amplicons; Acetobacteriaceae and Proteobacteria dominant organisms | [46] |
| American coolship ale beer fermentation | V4 16S rDNA | >150 | Illumina | Family/genus | Greengenes | Initial Enterobacteriaceae are overgrown by LAB species though fermentation; <i>Pediococcus</i> spp. becomes dominant after few weeks | [95] |
| Winery-associated microbiota before, during and after harvest | V4 16S and ITS1 from rDNA of bacteria and fungi | >150 | Illumina | Family/genus | Greengenes and UNITE | <i>Saccharomyces</i> colonized winery surfaces; microbial communities were dependent on the production context at each site, shaped by technological practices, processing stage and season | [106] |
| Cocoa bean fermentation | Total DNA | na | 454 FLX | Species | RDP and NCBI nr | Complex fermentation including bacteria (<i>Acetobacter pasteurianus</i> and <i>Lb. fermentum</i>) and yeasts <i>Hansenula uvarum</i> , <i>Hansenula opuntiae</i> and <i>Saccharomyces cerevisiae</i>) | [111] |
| Korean rice beer fermentation | V1-V3 16S rDNA and ITS1 from bacteria and fungi | >350 | 454 FLX | Genus/species | Genus | Proteobacteria are replaced by LAB species through fermentation; amylolytic yeasts drive saccharification; alcohol-producing <i>S. cerevisiae</i>) dominate at the end of the fermentation | [94] |
| “Cheonggukjang” fermentation | V1-V2 16S rDNA | >300 | 454 FLX | Genus | SILVA | <i>Bacillus</i> spp. dominates the fermentation, but the actual majority species was different in different samples | [86] |
| Korean soybean pastes | V1-V2 16S rDNA | >300 | 454 FLX | Genus | SILVA | High diversity in different brands; <i>Bacillus</i> spp. dominant in a majority of products; LAB species occasionally dominant | [87] |
| “Kochujang” fermentation | V1-V2 16S rDNA | >300 | 454 FLX | Genus | SILVA | <i>Bacillus</i> spp. dominates the fermentation, but the actual majority species was different in different samples | [88] |
| Ten kinds of “kimchi” | V1-V3 16S rDNA | >300 | 454 FLX | Genus/ | ExTaxon | Bacteria diversity and richness varied highly and depended on the type of | [89] |

| | | | | | | | |
|---|--|------------|----------|------------------------------|----------------------|--|-------|
| Table olives fermentation | V1-V3 16S rDNA and cDNA | >250 | 454FLX | species Genus/ species | database NCIB nr | “kimchi” Agreement for the DNA and RNA data; high level of halophilic bacteria at the beginning of fermentation; <i>Lactobacillus</i> spp. at the end | [38] |
| Ray and wheat sourdough fermentation | V1-V3 16S rDNA and cDNA | >300 | 454FLX | Genus/ species | RDP | Grain-associated bacteria do not progress in dough, except for <i>E. coli</i> ; <i>Weissella</i> , <i>Lactobacillus</i> and <i>Leuconostoc</i> dominate the fermentation | [49] |
| Fermentation of medieval sushi (“kaburazushi”) | V1-V2 16S rDNA | >300 | 454 FLX | Genus | NCIB nr | <i>Bacillus</i> and <i>Staphylococcus</i> are replaced by <i>Lactobacillus</i> through fermentation; <i>Lb. sakei</i> constitutes 80% of the microbiota. | [91] |
| Milk and dairy products | | | | | | | |
| Subclinical mastitic milk | Total DNA | na | 454 FLX | Species | SEED | Presence of <i>E. coli</i> , <i>Pseudomonas mendocina</i> , <i>Staphylococcus aureus</i> and <i>Klebsiella pneumoniae</i> and their phages | [107] |
| Healthy and mastitic milk | V1-V2 16S rDNA | 250-500 | 454 FLX | Genus | RDP | DNA sequences from recognized pathogens, from pathogens not associated with mastitis and from bacteria not known to be pathogens | [96] |
| Healthy and culture-negative mastitic milk | V1-V2 16S rDNA | >200 | 454 FLX | Genus | SILVA RDP | <i>Pseudomonas</i> , <i>Ralstonia</i> , <i>Psychrobacter</i> reads higher in milk from healthy udders; uncultured bacteria as causative agents in some mastitis | [68] |
| Pasteurized, cultured raw milk | V3 16S rDNA | >300 | 454 FLX | Genus/ species | RDP | <i>S. thermophilus</i> dominant organism; <i>Lb. delbrueckii</i> , <i>Lb. helveticus</i> , and <i>Enterococcus</i> spp. subdominant components | [48] |
| Irish kefir grain and associated kefir beverage | V4 16S rDNA | nr | 454 FLX | Genus/ species | RPD | <i>Lactobacillus</i> spp., as majority organisms including <i>Lb. kefiranofaciens</i> , <i>Lb. kefiri</i> , <i>Lb. parabuchneri</i> , and <i>Lb. helveticus</i> . | [97] |
| Brazilian kefir grains | V3 16S rDNA | >300 | 454 FLX | Genus/ species | RDP | Dominant species <i>Lb. kefiranofaciens</i> , <i>Lb. kefiri</i> , <i>Lb. parakefiri</i> ; <i>Acetobacter</i> spp. in only one sample | [47] |
| Danish raw milk cheeses made with starter cultures | V3-V4 16S rDNA | >200 na | 454 FLX | Genus | RDP | <i>Streptococcus</i> , <i>Lactococcus</i> and <i>Lactobacillus</i> constitute the largest groups; minor discrepancies between pyrosequencing and DGGE | [98] |
| Oscypeck traditional PDO Polish cheese | V5-V6 16S rDNA | >150 | 454 FLX | Genus | SILVA | <i>Lactococcus</i> spp. as dominant organism; presence of <i>Bifidobacterium</i> spp. and <i>Enhydrobacter</i> spp. Reads | [99] |
| Milk, whey starters and Mozzarella cheese | V1-V3 16S rDNA | >200 | 454 FLX | Genus/ species | RDP | <i>Acinetobacter</i> spp. and <i>Pseudomonas</i> spp. dominant in milk; <i>S. thermophilus</i> , <i>Lb. delbrueckii</i> and <i>Lb. helveticus</i> in cheese. | [102] |
| Latin-style cheeses | V1-V3 16S rDNA | >200 | 454 FLX | Genus/ species | RDP | High bacterial diversity in different brands; presence of high numbers of <i>Exiguobacterium</i> in one of the brands | [101] |
| Raw milk cheeses | V3-V4 16S rDNA and cDNA | nr | 454 FLX | Genus | RDP | Fate of starters and inoculated pathogens in cheese. <i>Listeria innocua</i> and <i>Stap. aureus</i> do not progress in cheese but <i>E. coli</i> does | [37] |
| Artisanal Irish cheeses and associated cheese rinds | V4 16S rDNA | >150 | 454 FLX | Genus | NCBI nr | Detection for the first time in cheese of <i>Arthrobacter</i> , <i>Brachybacterium</i> , <i>Faecalibacterium</i> , <i>Helcococcus</i> and <i>Prevotella</i> | [100] |
| Artisan American cheeses | V4 16S and ITS1 from rDNA and cDNA of bacteria | >150 | Illumina | Family/ genus | Greengenes and UNITE | <i>Debaryomyces</i> and <i>Lactococcus</i> as dominant fungi and bacteria, respectively; similar microbial communities occupy the same type surfaces; facility-specific microorganisms | [103] |

| | | | | | | | |
|---|---------------------------------|------------|-------------------|-------------------|------------------|---|-------|
| Industrial starter for Gouda cheese | and fungi Genomes, total DNA | na | 454 FLX, Illumina | Strain | NCBI nr | Metabolic complementation of starter components, <i>Lc. Lactis</i> and <i>Leuc. mesenteroides</i> ; starter stability by “kill-the-winner”-phage dynamics | [109] |
| Meat and fish products | | | | | | | |
| “Jeotgal” (fermented fish and seafood products) | V3 16S rRNA, archaea, bacteria | >100 | 454 FLX | Family | Greengenes | High microbial diversity, <i>Halorobum</i> , <i>Alkalalicoccus</i> , dominant archaeal genera; <i>Weissella</i> and <i>Lactobacillus</i> major LAB genera | [44] |
| Packaging of beef meat | V1-V3 16S rDNA | 500 | 454 FLX | Genus/ species | RDP | Inhibition of Enterobacteriaceae by modified atmospheres; nisin-active vacuum packaging inhibited <i>Brochrotys thermosphacta</i> | [105] |
| Marinated and unmarinated broiler meat | V1-V3 16S rDNA, total DNA | >250 na | 454 FLX | Family/ genus | SILVA NCBI nr | Carnobacteriaceae, Clostridiaceae, Enterococcaceae, Enterobacteriaceae, and Vibrionaceae more common in unmarinated meat samples | [110] |

^aRDP, Ribosomal Database Project; NCIB nr, National Center for Biotechnology Information non-redundant nucleotide database; MG-RAST, Meta Genome Rapid Annotation server based on Subsystem Technology; SILVA (quality-controlled database of aligned ribosomal RNA gene sequences); Greengenes (a chimera-checked 16S rRNA gene database); EzTaxon (a web-based tool for identification of prokaryotes based on 16S rDNA sequences).

nr, not reported.

na, not applicable.

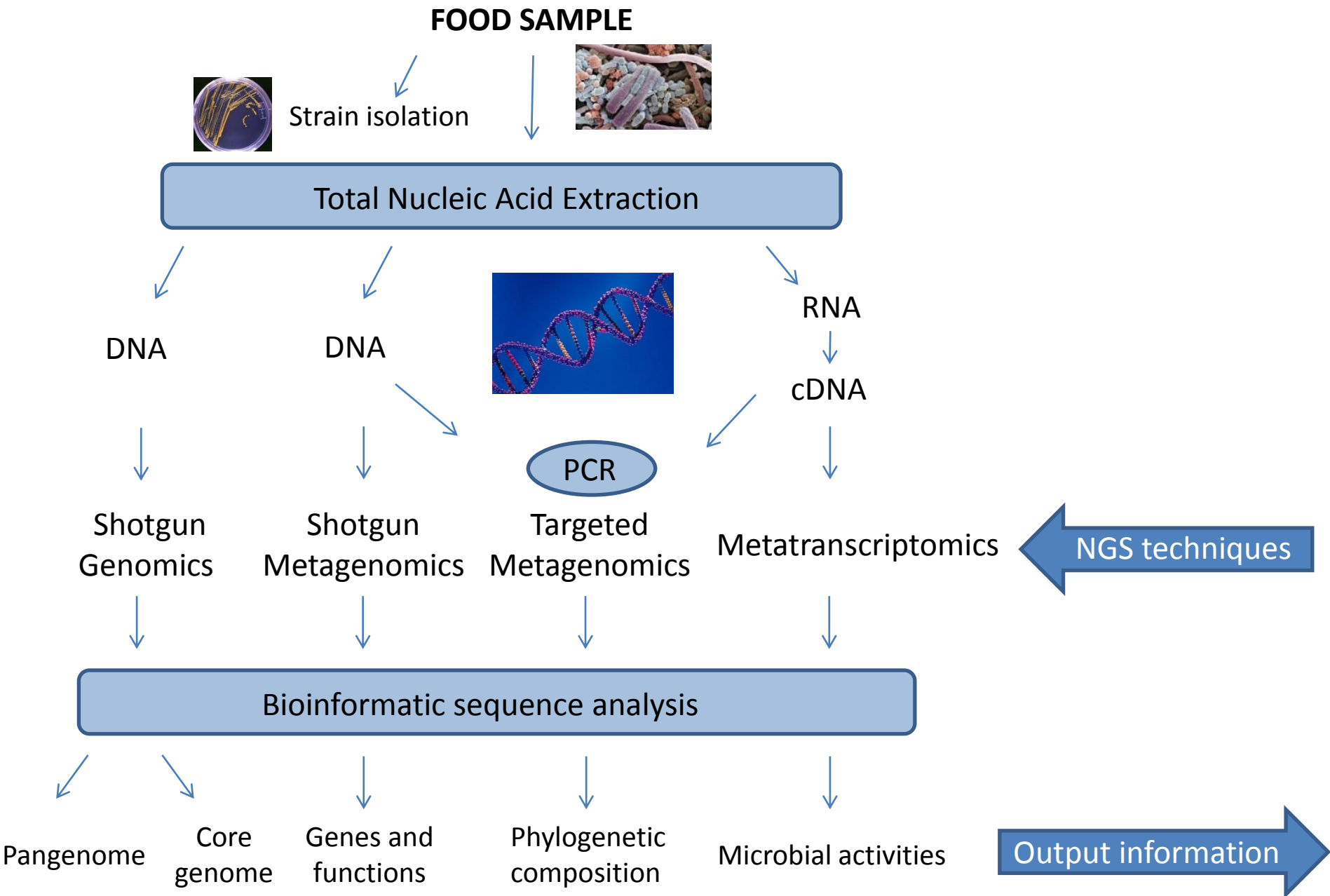


Figure 1.- Next-Generation Sequencing (NGS) applications in Food Microbiology and general flow chart.

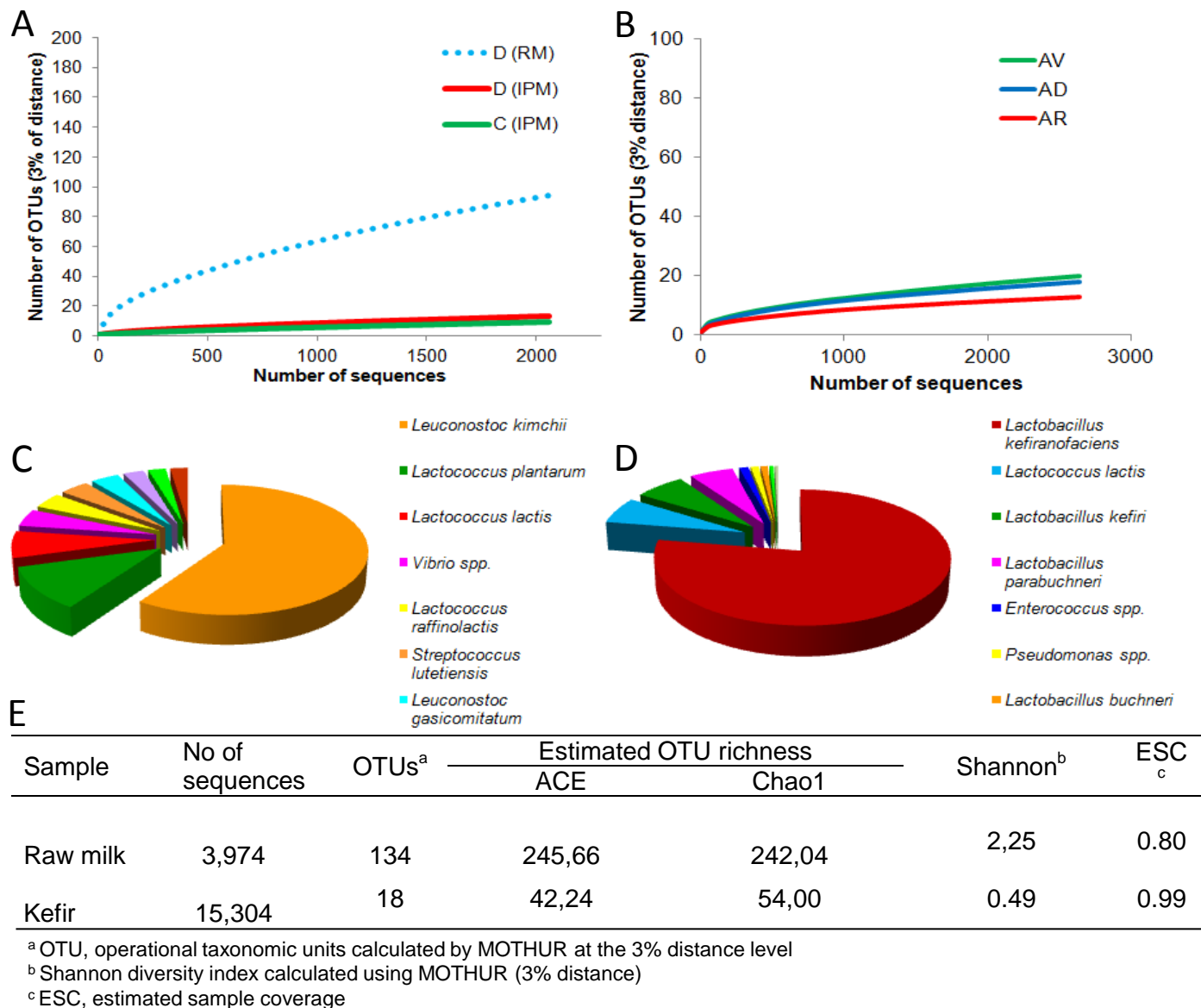


Figure 2.- Rarefaction curves of raw milk (dotted blue line) and pasteurized milk samples incubated at 42°C for 24 h (red and green lines (**Panel A**)). Rarefaction curves of several samples of Brazilian kefir grains (Panel B). Schematic representation of the microbiota from raw milk (Panel C) and a kefir (Panel D) samples. Estimated OTU richness, sample coverage and diversity indexes are indicated on the table (Panel E) (data from Leite et al., 2012 and Delgado et al., 2013).